Volume 7
Lectures in Applied Mathematics

SPACE MATHEMATICS
PART 3

J. Barkley Rosser, EDITOR
MATHEMATICS RESEARCH CENTER
THE UNIVERSITY OF WISCONSIN

1966
AMERICAN MATHEMATICAL SOCIETY, PROVIDENCE, RHODE ISLAND

Supported by the
National Aeronautics and Space Administration under Research Grant NsG 358
Air Force Office of Scientific Research under Grant AF-AFOSR 258-63
Army Research Office (Durham) under Contract DA-31-124-ARO(D)-82
Atomic Energy Commission under Contract AT(30-1)-3164
Office of Naval Research under Contract Nonr(G)00025-63
National Science Foundation under NSF Grant GE-2234

Library of Congress Catalog Card Number 66-20435

Printed in the United States of America

Contents

Lectures in Applied Mathematics

Proceedings of the Summer Seminar, Boulder, Colorado, 1960

VOLUME I — LECTURES IN STATISTICAL MECHANICS
G. E. Uhlenbeck and G. W. Ford with E. W. Montroll

VOLUME II — MATHEMATICAL PROBLEMS OF RELATIVISTIC PHYSICS
I. E. Segal with G. W. Mackey

VOLUME III — PERTURBATION OF SPECTRA IN HILBERT SPACE
K. O. Friedrichs

VOLUME IV — QUANTUM MECHANICS
V. Bargmann

Proceedings of the Summer Seminar, Ithaca, New York, 1963

VOLUME V — SPACE MATHEMATICS, PART 1
J. Barkley Rosser, Editor

VOLUME VI — SPACE MATHEMATICS, PART 2
J. Barkley Rosser, Editor

VOLUME VII — SPACE MATHEMATICS, PART 3
J. Barkley Rosser, Editor

S. F. Shen

Basic Fluid Dynamics

I. **Introduction.** In attempting to survey basic fluid dynamics in a program dedicated to the field of applied mathematics in space problems, the foremost question is to settle upon what should be meant by the word "basic". To this end, Professor Goldstein's admirable monograph [1] has provided a valuable guiding principle. Our endeavor in the following, however, is different from an abbreviated version of Goldstein's book, and reflects somewhat the aerodynamicist's view point. After the formulation of the general equations of motion, the emphasis is mostly on the motivation and derivation of the different approximations which find applications in various practical problems, particularly to bodies in flight at the higher speed ranges typical of space activities. Much material of basic and mathematical interest is unavoidably left out, as are the full details of the solution of any specific problem. In their places, we choose rather to illustrate, ever so briefly to be sure, how the theory has been exploited in the explanation and prediction of complicated physical phenomena.

Since most of the coverage is "basic", therefore contained in the well-known treatises such as those of Lamb [3] and Milne-Thomson [4], as well as Goldstein's book mentioned above, we have refrained from giving references except in rare instances.

II. **Description of fluid motion.** The fluid medium we work with shall be a continuum which, although somewhat idealized, should approximate the real gas of interest, namely air, in its behavior. Fluid dynamics then deals with such a gas in motion with or without the presence of solid boundaries. The state of gas in equilibrium, as when enclosed in a stationary and insulated vessel, is described by two thermodynamic variables, say density ρ and temperature T; and any other thermodynamic variable can be expressed in terms of ρ and T. In particular, we often desire to know the pressure p of the gas, observable as the normal force per unit area acting on the wall. The relation may be written as

$$p = p(\rho, T) \tag{II.1}$$

and is usually referred to as the "equation of state". Under the assumption of a perfect gas, Equation (II.1) becomes explicitly

$$p = \rho RT \tag{II.2}$$

where R is the gas constant, depending only upon the molecular weight of the gas.

When a body of gas is in arbitrary motion, it becomes necessary to regard the body of gas as composed of a large number of fluid elements, which must be small enough to represent the details of the fluid motion, yet not so small as to exhibit the coarse nature of the molecular motion. A velocity $\mathbf{V}$ may be assigned to each fluid element, and an observer riding with the fluid element may now determine the density and temperature of the gas in the fluid element. The pressure p follows again from Equation (II.1). If we trace the changes of p, ρ, T, $\mathbf{V}$ with time for each fluid element, the result is the "Lagrangian description" of the fluid motion. Alternatively, it is often more convenient for analysis to use a field representation by examining the flow pattern, i.e., the functions

$$p(\mathbf{r}, t), \rho(\mathbf{r}, t), T(\mathbf{r}, t), \mathbf{V}(\mathbf{r}, t)$$

where $\mathbf{r}$ designates the location of the fluid element at the given time t. This is now the "Eulerian description" of the fluid motion.

In the Eulerian description, the rate of change of any property Q of a *given* fluid element is usually written as DQ/Dt. Hence if Q is expressible as $Q(\mathbf{r}, t)$, we have

$$\frac{D}{Dt}Q = \lim_{\Delta t \to 0} [Q(\mathbf{r} + \Delta\mathbf{r}, t + \Delta t) - Q(\mathbf{r}, t)]/\Delta t$$

$$= \lim_{\Delta t \to 0} [Q(\mathbf{r} + \mathbf{V}\Delta t, t + \Delta t) - Q(\mathbf{r}, t)]/\Delta t \tag{II.3}$$

$$= \left(\frac{\partial}{\partial t} + \mathbf{V} \cdot \nabla\right) Q.$$

For example for given $\mathbf{V}(\mathbf{r}, t)$, the acceleration of the fluid element is equal to $D\mathbf{V}/Dt$. However, sometimes Q may not be given as a field, then a direct evaluation is necessary. To illustrate the latter, let Q be the volume $\delta\tau$ of a fluid element, and define the "dilatation" θ as the rate of volume change of the fluid element, per unit volume:

$$\theta \equiv \lim_{\delta\tau \to 0} \frac{1}{\delta\tau}\left(\frac{D}{Dt}\delta\tau\right). \tag{II.4}$$

If $\delta\tau$ is bounded by surface S and $\mathbf{n}$ is the outward unit normal on the surface element dS (See Figure 1), clearly by definition

$$\theta \equiv \lim_{\delta\tau \to 0} \int_S \mathbf{V} \cdot \mathbf{n} dS/\delta\tau$$

$$= \operatorname{div} \mathbf{V}. \tag{II.5}$$

Of considerable interest in the Eulerian description of fluid motion is the "streamline pattern", showing the direction of motion of each fluid element at a given instant. The "streamlines" are defined by

$$d\mathbf{r}_s \times \mathbf{V} = 0, \tag{II.6}$$

where $d\mathbf{r}_s$ is a length element on the streamline. If the flow pattern does not vary with time, the fluid motion is said to be a "steady flow". The streamlines in such cases coincide with the trajectories of the fluid elements.

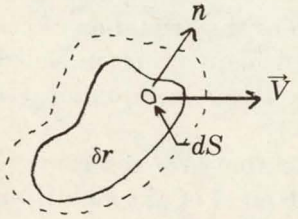


FIGURE 1. Dilation.

Aside from the translational motion of the fluid element, we must, of course, also expect in general a rotational motion as well as a change of shape with time. The angular velocity of the fluid element turns out to be one half the "vorticity", ω, which is defined through a given velocity field $\mathbf{V}(\mathbf{r}, t)$ as

$$\omega \equiv \nabla \times \mathbf{V}. \tag{II.7}$$

Following Equation (II.6), we may then look at the vorticity pattern by introducing "vortex lines" analogous to the streamlines,

$$d\mathbf{r}_v \times \omega = 0 \tag{II.8}$$

where $d\mathbf{r}_v$ is a length element on the vortex line.

III. **Equations of fluid motion.** The equations of fluid motion express the requirements that the fundamental laws of the conservation of mass, momentum and energy must not be violated. These can be very simply stated if the Lagrangian description is adopted. Consider a small fluid element of volume $\delta\tau$; its mass will be $\rho\delta\tau$. For generality we introduce a "mass source" $\dot{m}$ such that mass is being added to the fluid element at the rate of $\dot{m}\delta\tau$. Then the law of conservation of mass as applied to $\delta\tau$ states that

$$\frac{D}{Dt}\rho\delta\tau = \dot{m}\delta\tau. \tag{III.1}$$

In Lagrangian sense, the left-hand side is an ordinary time derivative of a product, and we may write

$$\delta\tau\frac{D\rho}{Dt} + \rho\frac{D}{Dt}\delta\tau = \dot{m}\delta\tau.$$

By using the definition of the dilatation θ, Equation (II.4), to evaluate $D\delta\tau/Dt$, the result may be rewritten as

$$\frac{D\rho}{Dt} + \rho\theta = \dot{m} \tag{III.2}$$

which is known as the "equation of continuity". When ρ, $\mathbf{V}$ and $\dot{m}$ are regarded as field quantities in Eulerian description, we only need to interpret the terms in Equation (III.2) according to Equations (II.3) and (II.5).

For the momentum $\rho\mathbf{V}\delta\tau$ and energy $\rho E\delta\tau$ (E, being defined as the energy per unit mass of the fluid), equations similar to Equation (III.1) may be written with a "momentum source" $\mathbf{P}$ and an "energy

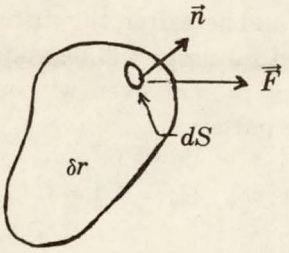


FIGURE 2. Stress.

source" E, respectively, on the right-hand side. The same manipulation yields

$$\rho \frac{D\mathbf{V}}{Dt} = \mathbf{P} - \dot{m}\mathbf{V} \tag{III.3}$$

and

$$\rho \frac{DE}{Dt} = \dot{E} - \dot{m}E. \tag{III.4}$$

Again, although derived from the Lagrangian description, Equations (III.3) and (III.4) offer no difficulty in interpretation for the Eulerian description, provided $\dot{m}$, $\mathbf{P}$ and $\dot{E}$ are given as field quantities.

We restrict ourselves in the following to the case of $\dot{m} = 0$. To proceed further with the terms $\mathbf{P}$ and $\dot{E}$, it will be assumed that these are only due to the interactions between adjacent fluid elements, and that the basic fluid properties are isotropic, namely, invariant with orientation. For $\mathbf{P}$, experience shows that any nonuniformity of motion causing a change of shape of the fluid element would be resisted by the fluid through the development of internal stresses between fluid elements. For $\dot{E}$, experience shows that heat will flow through the boundary of the fluid element if a nonuniformity of temperature exists. In addition, the stresses acting on the boundary perform mechanical work on the fluid element.

Consider now a fluid element $\delta\tau$ within the surface S. On a surface element dS, let $\mathbf{n}$ be the unit outward normal and $\mathbf{F}$ the resultant stress vector (see Figure 2). Referring to a set of Cartesian coordinates x_i, $i = 1, 2, 3$, these have components n_i and F_i, respectively. It is then convenient to introduce a stress tensor τ_{ij} such that

$$F_i = \tau_{ij} n_j. \tag{III.5}$$

In Equation (III.5) and hereafter, the customary convention of summing over a repeated subscript will be understood. Since τ_{ij} includes the pressure which is present even without fluid motion, we may separate τ_{ii} into two parts:

$$\tau_{ij} = \tau'_{ij} - p\delta_{ij} \quad (\delta_{ij} = 0, i \neq j; \quad \delta_{ij} = 1, i = j), \tag{III.6}$$

the negative sign indicating that the pressure is always opposite to $\mathbf{n}$. The tensor τ'_{ij} is the "viscous part" of τ_{ij}, and remains to be related to the nonuniformity of the fluid motion.

If the nonuniformity of the fluid motion is slight, it may be characterized by the first derivatives of $\mathbf{V}$ with respect to the space variables at the point under consideration, hence the tensor $\partial u_i/\partial x_j$. Splitting $\partial u_i/\partial x_j$ into symmetrical and anti-symmetrical parts, we have

$$\frac{\partial u_i}{\partial x_j} = e_{ij} + \omega_{ij},$$

$$e_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right), \tag{III.7}$$

$$\omega_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} - \frac{\partial u_j}{\partial x_i}\right).$$

The anti-symmetrical part ω_{ij} is easily seen to be

$$\omega_{ij} = -2\epsilon_{ijk}\omega_k \tag{III.8}$$

where ω_k is the component of the vorticity $\boldsymbol{\omega}$ defined by Equation (II.7) and ϵ_{ijk} is the alternating symbol,

$\epsilon_{ijk} = 0$ when the subscripts are not all different;

$= 1$ when i, j, k follow the cyclic order 1, 2, 3;

$= -1$ when i, j, k do not follow the cyclic order 1, 2, 3.

Thus ω_{ij} represents the nonuniformity due to a rigid rotation of the fluid. The change of shape of the fluid element as it moves along is entirely represented by the symmetrical tensor e_{ij}. To proceed further, the "viscous hypothesis" is made that τ'_{ij} should be linearly proportional to e_{ij}, i.e.,

$$\tau'_{ij} = C_{ijkl}e_{kl},$$

C_{ijkl} being constants. Now the physical law must not be affected by the orientation of x_1, x_2, x_3 in an isotropic fluid. Then (see [2, p. 70]) there must be

$$C_{ijkl} = \lambda\delta_{ij}\delta_{kl} + \mu(\delta_{ik}\delta_{jl} + \delta_{jk}\delta_{il})$$

where λ and μ are constants. Hence

$$\tau'_{ij} = \lambda\delta_{ij}e_{kk} + 2\mu e_{ij} \tag{III.9}$$

where obviously $e_{kk} = \nabla \cdot \mathbf{V}$.

There are then normal viscous stresses τ'_{11}, τ'_{22}, τ'_{33}. Summing the three, we have

$$\tau'_{ii} = (3\lambda + 2\mu)e_{ii}.$$

Thus, like pressure p, the average of the normal viscous stresses is independent of the axes. It is however proportional to the dilatation, and the coefficient $3\lambda + 2\mu$ is referred to as the "bulk viscosity coefficient". For monatomic gases, kinetic theory shows that

$$3\lambda + 2\mu = 0 \quad \text{or} \quad \lambda = -\frac{2}{3}\mu.$$

This result is generally assumed in most applications involving air (even though it is composed of primarily diatomic gases) so that the viscous stresses are all proportional to a single material constant, the "viscosity coefficient". Equation (III.9) becomes

$$\tau'_{ij} = -\frac{2}{3}\mu\delta_{ij}e_{kk} + 2\mu e_{ij} \tag{III.9$'$}$$

known as the "Navier-Stokes relation". It may be noted here that the viscosity coefficient is mainly a function of the temperature T.

We next turn to the heat flux due to the nonuniformity of the temperature field. Since T is a scalar, the nonuniformity is characterized by a vector ∇T. If $\dot{\mathbf{q}}$ is the heat flux vector (the rate of heat flow per unit area), the assumption of linear dependence leads in an analogous manner to

$$\dot{\mathbf{q}} = -k \nabla T \tag{III.10}$$

where the proportionality constant k is the "coefficient of thermal conductivity". Equation (III.10) is known as the "Fourier law". From a molecular viewpoint, both μ and k owe their origin to the random motion of the molecules, and these two are closely related.

For example, for monatomic gases, kinetic theory predicts

$$2k = 5\mu C_v$$

where C_v is the specific heat at constant volume.

With Equations (III.9′) and (III.10), it is now possible to represent **P** and $\dot{E}$ explicitly. For **P**, there is

$$P_i \delta\tau = \int_S F_i dS$$

$$= \int_S \tau_{ij} n_j dS$$

$$= \int_{\delta\tau} \frac{\partial}{\partial x_j} \tau_{ij} d\tau,$$

by Gauss' Theorem: or, as $\delta\tau \to 0$,

$$P_i = \frac{\partial}{\partial x_j} \tau_{ij} = -\frac{\partial p}{\partial x_i} + \frac{\partial}{\partial x_j} \tau'_{ij}. \tag{III.11}$$

For $\dot{E}$, there is

$$\dot{E}\delta\tau = \int_S \boldsymbol{F} \cdot V dS + \int_S \mathbf{n} \cdot k \nabla T dS$$

$$= \int_S \left(\tau_{ij} u_i + k \frac{\partial T}{\partial x_j} \right) n_j dS$$

$$= \int_{\delta\tau} \frac{\partial}{\partial x_j} \left[\tau_{ij} u_i + k \frac{\partial T}{\partial x_j} \right] d\tau$$

where u_i is the component of **V** in the x_i-direction. As $\delta\tau \to 0$, it follows that

$$\dot{E} = \frac{\partial}{\partial x_j} (\tau_{ij} u_i) + \frac{\partial}{\partial x_j} \left(k \frac{\partial T}{\partial x_j} \right). \tag{III.12}$$

Hence, with Equation (III.11) the momentum equation, Equation (III.3), becomes finally

$$\rho \frac{Du_i}{Dt} = -\frac{\partial p}{\partial x_i} + \frac{\partial}{\partial x_j} \tau'_{ij}. \tag{III.13}$$

In the energy equation, Equation (III.4), we note that for the fluid

element in motion,

$$E = U + \frac{1}{2} V^2$$

where U is the internal energy of the fluid element. Together with Equation (III.12), Equation (III.4) after simple manipulation becomes finally

$$\rho \frac{DU}{DT} = - pe_{jj} + \tau'_{ij} e_{ij} + \frac{\partial}{\partial x_j}\left(k \frac{\partial T}{\partial x_j}\right). \tag{III.14}$$

The second term of the right-hand side clearly represents the work done by the viscous stresses, and often is defined as the "dissipation function" Φ. It may be easily verified that $\Phi \geqq 0$ when τ'_{ij} is given by Equation (III.9′).

Alternative forms of the energy equation, Equation (III.14), are sometimes useful. For instance, in terms of the entropy S, since by thermodynamic definition

$$TdS = dU + pd\left(\frac{1}{\rho}\right),$$

the continuity equation, Equation (III.2) (with $\dot{m} = 0$), and Equation (III.14) combined lead to

$$\rho T \frac{DS}{Dt} = \Phi + \frac{\partial}{\partial x_j}\left(k \frac{\partial T}{\partial x_j}\right). \tag{III.14′}$$

In terms of the enthalpy h, since by thermodynamic definition

$$h = U + \frac{p}{\rho},$$

Equation (III.14) may also be replaced by

$$\rho \frac{Dh}{Dt} = \frac{Dp}{Dt} + \Phi + \frac{\partial}{\partial x_j}\left(k \frac{\partial T}{\partial x_j}\right). \tag{III.14″}$$

IV. **Physical boundary conditions of fluid motion.** The fluid motion has been defined in the above through the unknowns p, ρ, T and $\mathbf{V}$, which are required to satisfy Equations (II.2), (III.2) (with $\dot{m} = 0$), (III.13) and (III.14). A typical problem is to find the resultant fluid motion when an obstacle moves through the fluid in a prescribed manner. In the fluid domain, there remains the question of relating

the values of these unknowns for the fluid elements in contact with the obstacle with the prescribed motion and properties of the obstacle itself. We may think of Equations (II.2) and (III.2) as defining p and ρ in terms of $\mathbf{V}$ and T, so Equations (III.13) and (III.14) are really the equation to be integrated. Thus, if the obstacle is impermeable and represented by the surface $F_S(x_i, t) = 0$ and its temperature by the condition $T = T_S(t)$ on $F_S = 0$, we are interested to assign values of $\mathbf{V}$ and T for the fluid elements satisfying $F_S = 0$.

Now the resultant velocity of a point on the obstacle must satisfy $DF_S/Dt = 0$. Since the obstacle is assumed to be impermeable, the velocity of the fluid element at the same point must have the same velocity component normal to the surface, and therefore satisfy also $DF_S/Dt = 0$, although the tangential velocity is still arbitrary. We refer to this as the "condition of no penetration", or

$$\frac{DF_S}{Dt} = 0 \text{ for fluid elements on } F_S = 0. \tag{IV.1}$$

Obviously by the same reasoning, Equation (IV.1) is also the condition at the interface between two dissimilar fluids.

As for the tangential component of the fluid velocity and the temperature of the fluid element at the boundary, one usually appeals to experience whenever the mathematical solution requires these data. Ordinarily it is *assumed* that the fluid element shall have neither a relative velocity with respect to the boundary—the "condition of no slip", nor any temperature differences from that of the boundary—the "condition of no jump". These are confirmed as first approximations by kinetic theory considerations, so long as the gas is not too rarified.

The precise conditions under which the mathematical problem will be "properly set" is in general a difficult question because of the complicated nonlinear nature of the equations. The practice is rather to look for a solution when physically the problem is well-defined and can be set up in an experimental investigation. However, it should be noted that empirically under seemingly indentical conditions the observed flow may be either "laminar" or "turbulent". Take the steady flow through a circular pipe as an example: A mathematical solution of the equations predicts that the flow should move in layers; this is indeed well confirmed experimentally, but only if the flow velocity is relatively small. At higher velocities, the actual flow is composed of a steady mean motion superposed by

random time-dependent fluctuations. This phenomenon is typical rather than exceptional. It strongly suggests that the general uniqueness condition for the system of equations describing fluid flow would be extremely difficult to lay down.

V. **Rotational and irrotational motions.** We defined in Equations (II.7) the vorticity vector ω

$$\omega = \nabla \times \mathbf{V}$$

representing the rigid body rotation of the fluid element. Kinematically, the fluid motion may be classifled as rotational or irrotational depending on whether $\omega \neq 0$ in general or $\omega = 0$ everywhere in the fluid. We note first that because ω is the curl of $\mathbf{V}$, it is a solenoidal vector, i.e., $\nabla \cdot \omega = 0$. Now if we have a steady flow field with $\dot{m} = 0$, the equation of continuity, Equation (III.2), reduces to $\nabla \cdot (\rho \mathbf{V}) = 0$. Thus by analogy it can be said that ω also satisfies an "equation of continuity". Let a contour C enclosing a surface S be drawn in the fluid (see Figure 3). Vortex lines can be passed through points of C to form a "vortex tube", and the following must hold:

$$\int_S \omega_i \, dS_i = \int_{S'} \omega_i \, dS_i$$

where S' is the area enclosed by C' anywhere downstream along the vortex lines from C. In particular, by taking $S \to 0$, the vortex tube becomes a very thin "vortex element". Thus a vortex element can never end within the fluid. It may, however, form a closed loop.

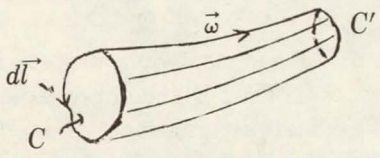


FIGURE 3. Vortex Tube.

If the fluid motion is an irrotational motion, by applying Stokes' theorem to the contour C,

$$\oint_C \nabla \cdot d\mathbf{l} = \int_S \omega_i \, dS_i = 0.$$

Consequently a "velocity potential" ϕ exists such that

$$\mathbf{V} = \nabla \phi.$$

Since a scalar ϕ defines the velocity vector, the mathematical problem is then to find the solution for a single function ϕ and becomes much simplified. It is therefore of interest to examine the circumstances under which the irrotational approximation may be adopted.

With $\dot{m} = 0$, consider the momentum equation, Equation (III.3),

$$\frac{D\mathbf{V}}{Dt} = \frac{1}{\rho}\mathbf{P}.$$

Expanding $D\mathbf{V}/Dt$, we have

$$\frac{D\mathbf{V}}{Dt} = \frac{\partial \mathbf{V}}{\partial t} + \frac{1}{2}\nabla \mathbf{V}^2 - \mathbf{V} \times \omega. \tag{V.1}$$

Hence, by straightforward manipulation and with Equation (III.2),

$$\begin{aligned}\nabla \times \frac{D\mathbf{V}}{Dt} &= \frac{\partial}{\partial t}\nabla \times \mathbf{V} - \nabla \times (\mathbf{V} \times \omega) \\ &= \frac{D\omega}{Dt} - (\omega \cdot \nabla)\mathbf{V} - \frac{\omega}{\rho}\frac{D\rho}{Dt} \\ &= \rho\left[\frac{D}{Dt}\left(\frac{\omega}{\rho}\right) - \left(\frac{\omega}{\rho} \cdot \nabla\right)\mathbf{V}\right].\end{aligned}$$

The "vorticity equation" follows immediately

$$\frac{D}{Dt}\left(\frac{\omega}{\rho}\right) = \left(\frac{\omega}{\rho} \cdot \nabla\right)\mathbf{V} + \frac{1}{\rho}\nabla \times \left(\frac{1}{\rho}\mathbf{P}\right). \tag{V.2}$$

We shall examine the behavior of ω under the following simplifying conditions:

(1) $p = p(\rho)$, e.g., $p \propto \rho^{\gamma}$ for isentropic process (γ being the ratio of specific heats), or $\rho = \text{const}$ for incompressible fluid;

(2) $\mu = 0$, the inviscid approximation. Under the simplification, Equation (III.11) gives $\mathbf{P} = -\nabla p$, and since $\nabla \times ((1/\rho)\nabla p) = \nabla \times (\nabla \int (dp/\rho)) = 0$, Equation (V.2) becomes

$$\frac{D}{Dt}\frac{\omega}{\rho} = \left(\frac{\omega}{\rho} \cdot \nabla\right)\mathbf{V}$$

or

$$\begin{aligned}\frac{D}{Dt}\frac{\omega_j}{\rho} &= \frac{\omega_i}{\rho}\frac{\partial}{\partial x_j}u_i \\ &= \frac{\omega_j}{\rho}\left[e_{ij} + \omega_{ij}\right].\end{aligned}$$

Noting Equation (III.8),

$$\omega_j \omega_{ij} = -2\epsilon_{ijk}\omega_j \omega_k = 0.$$

We finally get

$$\frac{D}{Dt}\frac{\omega_i}{\rho} = \frac{\omega_j}{\rho} e_{ij} \tag{V.3}$$

which is sometimes interpreted as saying that following the fluid element, ω_i/ρ changes due to the "stretching" of the vortices. In particular, for two-dimensional motion $\boldsymbol{\omega} = (0, 0, \omega_3)$ but $e_{33} = 0$, so that Equation (V.3) degenerates into

$$\frac{D}{Dt}\frac{\omega_3}{\rho} = 0 \tag{V.4}$$

saying that the vorticity, strictly speaking ω_3/ρ, is attached to the fluid element without change. Following Equation (V.4), as long as $p = p(\rho)$ *and* in the inviscid limit, if at some time the fluid element does not possess vorticity it will not acquire vorticity in two-dimensional motion. When the flow field is set up from rest through the arbitrary movements of a two-dimensional body, we therefore expect irrotational motion at all times. For the general three-dimensional flow, the same conclusion can be reached by integrating Equation (V.3) for a given fluid element (see [4, p. 84]). These are of course only useful in practical cases when the underlying assumptions are acceptable.

Let us now examine the role of viscosity. Consider for simplicity the small perturbation from a state of rest, i.e., $\mathbf{V} = \mathbf{V}'$, $\boldsymbol{\omega} = \boldsymbol{\omega}'$, $\rho = \rho_0 + \rho'$, etc., ρ_0 being the density of the fluid at rest and primed quantities being the small perturbations. After neglecting the quadratic terms involving the perturbation quantities and with the help of Equations (III.11) and (III.9′), Equation (V.2) is reduced to the diffusion equation

$$\frac{\partial}{\partial t}\boldsymbol{\omega}' = \nu_0 \nabla^2 \boldsymbol{\omega}' \tag{V.5}$$

where $\nu_0 = \mu_0/\rho_0$, the kinematic viscosity. If a vortex element is generated in an infinite fluid at $t = 0$ and maintained afterwards, the consequence of Equation (V.5) is that the vorticity will spread out, with decreasing strength, to occupy a region of size $(\nu_0 t)^{1/2}$ beyond which the effect is *essentially* nil. This result is *qualitatively* useful in

visualizing the flow patterns surrounding a body moving in a fluid at rest. Suppose a thin two-dimensional plate of length L moves parallel to itself in a viscous fluid at constant velocity $\mathbf{V}$ (see Figure 4). An obvious irrotational solution is that the fluid is undisturbed, satisfying all differential equations except for the viscous "no-slip" and "no-jump" conditions at the surface. We imagine viscosity to be absent for $t < 0$, but suddenly turned on at $t = 0$. The fluid elements in contact with the plate will be instantaneously arrested, creating a surface of discontinuity which may be interpreted as a vortex sheet composed of concentrated vortex elements. The vorticity subsequently spreads out approximately at a rate $O(\sqrt{(\nu_0/t)})$.

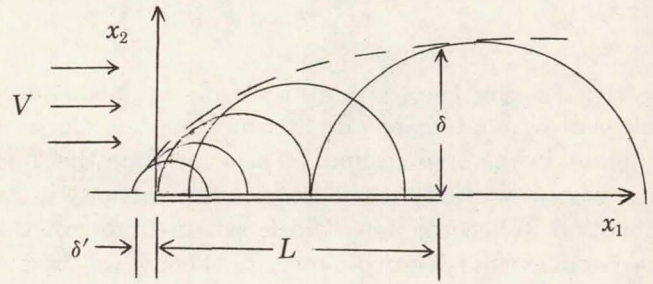


FIGURE 4. Moving Plate.

To an observer fixed relative to the plate, the instantaneous flow pattern will be swept downstream at a speed equal to V and the vorticity will be seen as essentially confined in a region roughly parabolic starting from the leading edge of the plate, the thickness δ reaching a value $O(\sqrt{(\nu_0 L/V)})$ at the end of the plate. In nondimensional form, we have therefore

$$\frac{\delta}{L} \sim O(1/\sqrt{(\mathrm{Re})})$$

where $\mathrm{Re} = VL/\nu_0$, the "Reynolds number" based on the length L. There would be furthermore a disturbed region ahead of the plate of size δ', given by

$$\delta' \sim O(\sqrt{(\nu_0 \delta'/V)})$$

or, again in terms of a Reynolds number, $\mathrm{Re}_{\delta'} \equiv V\delta'/\nu_0 \sim O(1)$. Thus the size of the region of rotational flow because of the viscous effects is confined to the immediate neighborhood of the plate as $\nu_0 \to 0$. In

fact, the thickness δ' tends to zero much faster than the thickness δ. The layer of $O(\delta)$ adjacent to the body is referred to as the "boundary layer". The viscous and rotational region swept behind the body is the "wake". Outside of the thin boundary layer and the wake, the flow is seen to be essentially irrotational. For blunt bodies these qualitative descriptions remain valid, but although the boundary layer thickness is still proportional to $\sqrt{\nu_0}$, the wake will be of the order of the body thickness (see Figure 5). It should be mentioned that rotationality may also be present due to curved shock waves which form ahead of the body when it moves at high speeds. (See §X). This is an example where $p = p(\rho)$ is not true.

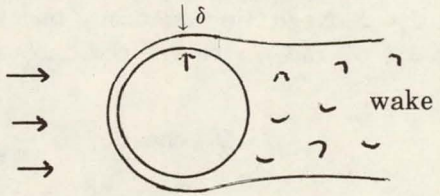


FIGURE 5. Moving Blunt Body.

Finally we note that the boundary layer and the wake are actually the corrections to an inviscid solution due to the viscous boundary conditions. Consequently outside of these regions whether the flow be rotational or irrotational, the fluid may be regarded as inviscid. As the kinematic viscosity of gases is usually very small, in most flow problems the Reynolds number will be large and the boundary layer will be relatively thin. Then the inviscid "no penetration" condition may be applied without serious error as if the boundary layer were absent. The difficulty of the unknown boundary of the wake, however, cannot be circumvented in constructing an inviscid approximation for blunt bodies.

VI. **The inviscid approximation.** Let us now exploit the inviscid approximation. Since the viscosity μ and the thermal conductivity k are of the same mechanism, the fluid should also be regarded as nonheat-conducting in the same approximation. The immediate consequence from Equation (III.14′) is

$$\frac{DS}{Dt} = 0$$

i.e., the entropy is constant following each fluid element, though not necessarily throughout the flow field. The "Navier-Stokes' equations", Equation (III.13), degenerate into the "Euler equations"

$$\rho \frac{D\mathbf{V}}{Dt} = -\nabla p. \tag{VI.2}$$

The continuity equation, Equation (III.2), of course is unaffected:

$$\frac{D\rho}{Dt} + \rho \nabla \cdot \mathbf{V} = 0. \tag{VI.3}$$

Consider again a small perturbation of the fluid from rest at pressure p_0 and density ρ_0. Neglecting quadratic terms of the perturbation quantities $\mathbf{V}'$, p' and ρ', we get the "acoustic theory" from Equations (VI.1—3):

$$\begin{aligned} &S = S_0, \text{ const.}, \\ &\rho_0 \frac{\partial \mathbf{V}}{\partial t} = -\nabla p', \\ &\frac{1}{\rho_0} \frac{\partial \rho'}{\partial t} + \nabla \cdot \mathbf{V}' = 0. \end{aligned} \tag{VI.4}$$

The first of these may alternatively be expressed as

$$\begin{aligned} &p/p_0 = (\rho/\rho_0)^\gamma, \quad \gamma = C_p/C_v, \\ &\text{or} \\ &p' = a^2 \rho' \end{aligned} \tag{VI.5}$$

where $a = \sqrt{(\partial p/\partial \rho)_S} = \sqrt{(\gamma p_0/\rho_0)}$, the "speed of sound". Eliminating p' and $\mathbf{V}'$ in favor of ρ', we find

$$\frac{\partial^2}{\partial t^2} \rho' - a^2 \nabla^2 \rho' = 0 \tag{VI.6}$$

which is the "wave equation". The elementary solution for introducing a small disturbance at a point at $t = 0$ is such that the disturbance spreads out in space at a rate equal to the sound speed "a", and beyond a radius of at the fluid is undisturbed. For a source of disturbance moving at constant velocity $\mathbf{V}$, to an observer fixed to the source of disturbance, two different flow patterns result depending on the "Mach number" $M \equiv V/a$ (see Figure 6). For $M < 1$ the disturbance spreads out in all directions, eventually swallowing up the entire space in a long enough time. For $M > 1$, the disturbed

region is conical, formed by the envelope to the drifting spheres, with the vertex at the source of disturbance, the semi-angle being equal to the "Mach angle" $\sin^{-1}1/M$. The conical surface itself for finite time, in conjunction with the spherical surface in the back, is the wave front separating the undisturbed and the disturbed regions.

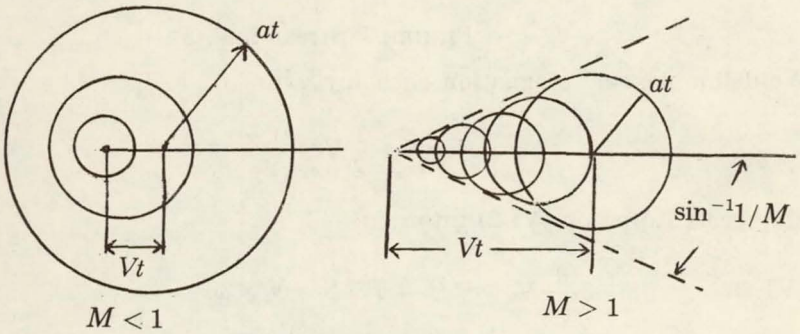


FIGURE 6. Acoustical Flow Patterns.

Following von Kármán, one may refer to the undistrubed region ahead of the conical surfact as the "zone of silence", and the disturbed region behind as the "zone of action". While the above is based upon linearized small perturbation theory, the difference in behavior of subsonic and supersonic flows remains qualitatively the same even if the disturbances caused by the moving object are no longer small.

Without restricting ourselves to small disturbances, we return to Equations (VI.1) to (VI.3). In certain cases, a first integral of the Euler equations, Equation (VI.2), can be directly obtained, which will further simplify the problem of finding a solution. By Equation (V.1), Equation (VI.2) may be written as

$$\frac{\partial \mathbf{V}}{\partial t} + \frac{1}{2}\nabla \mathbf{V}^2 - \mathbf{V} \times \boldsymbol{\omega} = -\frac{1}{\rho}\nabla p.$$

Now the definition of entropy S is, for a given element, $TdS = dh - (1/\rho)dp$. But inasmuch as T and ρ are always expressible as functions of p and h, this expression may also be regarded as an ordinary differential relation defining $S(p,h)$, hence leading to

$$T\nabla S = \nabla h - \frac{1}{\rho}\nabla p.$$

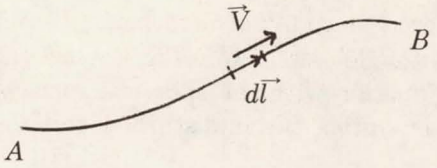


FIGURE 7. Streamline.

We define next a "stagnation enthalpy" H,

$$H \equiv h + \frac{1}{2}\mathbf{V}^2$$

and recast Equation (VI.2) into

$$\frac{\partial}{\partial t}\mathbf{V} + \nabla H - T\nabla S = \mathbf{V} \times \boldsymbol{\omega}. \tag{VI.7}$$

For the special case of steady flow, $(\partial/\partial t)\mathbf{V} = 0$, Equation (VI.7) is known as "Crocco's theorem", showing for instance that vorticity would arise due to entropy gradient. If Equation (VI.7) is dotted into the length element $d\mathbf{l}$ along a streamline, at given t (see Figure 7), and then integrated between the end points A and B, the result is

$$\frac{\partial}{\partial t}\int_A^B \mathbf{V} \cdot d\mathbf{l} + \int_A^B dH - \int_A^B TdS = 0.$$

This yields the so-called "Bernoulli's equation" in the following cases:

(a) For steady flow $(\partial/\partial t = 0)$, $DS/Dt = \mathbf{V} \cdot \nabla S = 0$, hence $\mathbf{V}$ and $d\mathbf{l}$ are both normal to ∇S. Consequently

$$H = \text{const. along any streamline.} \tag{VI.8}$$

(b) For irrotational flow with uniform entropy everywhere ("homentropic"). $\mathbf{V} = \nabla\phi$ and $\nabla S = 0$, hence

$$\frac{\partial\phi}{\partial t} + H = \text{const. along any streamline.} \tag{VI.9}$$

When the streamlines can always be traced to a region of steady uniform flow, the constant in the right-hand side of Equation (VI.8) or (VI.9) becomes identical for all points in the flow field.

We next proceed to derive the equation for the velocity potential

ϕ in an irrotational homentropic flow. The scalar product of $\mathbf{V}$ with Equation (IV.2) leads to

$$\frac{\partial}{\partial t}\frac{V^2}{2} + \frac{1}{2}(\mathbf{V}\cdot\nabla)V^2 = -\frac{a^2}{\rho}\mathbf{V}\cdot\nabla\rho$$

$$= \frac{a^2}{\rho}\left[\frac{\partial\rho}{\partial t} + \rho\nabla\cdot\mathbf{V}\right],$$

by Equation (VI.3). But differentiation of Equation (VI.9) gives

$$\frac{a^2}{\rho}\frac{\partial\rho}{\partial t} + \frac{\partial}{\partial t}\frac{V^2}{2} + \frac{\partial^2\phi}{\partial t^2} = 0.$$

Eliminating $\partial\rho/\partial t$ between the two expressions, we get

$$\text{(VI.10)} \qquad \frac{\partial}{\partial t}V^2 + \frac{\partial^2\phi}{\partial t^2} - a^2\nabla\cdot\mathbf{V} + \mathbf{V}\cdot\nabla\frac{V^2}{2} = 0.$$

In Cartesian coordinates, Equation (VI.10) may be written as

$$\phi_{tt} - (a^2 - \phi_x^2)\phi_{xx} - (a^2 - \phi_y^2)\phi_{yy} - (a^2 - \phi_z^2)\phi_{zz}$$

$$\text{(VI.10}') \qquad + 2(\phi_x\phi_y\phi_{xy} + \phi_x\phi_z\phi_{xz} + \phi_y\phi_z\phi_{yz}$$

$$+ \phi_x\phi_{xt} + \phi_y\phi_{yt} + \phi_z\phi_{zt}) = 0.$$

Here a^2 is expressible also in ϕ by noting that

$$\text{(VI.11}') \qquad H = \frac{\gamma}{\gamma-1}\frac{p}{\rho} + \frac{V^2}{2} = \frac{a^2}{\gamma-1} + \frac{V^2}{2}$$

while Equation (VI.9) shows that H is directly related to $\partial\phi/\partial t$. It is, however, more instructive without explicitly evaluating a^2. To fix ideas, suppose we have a body of characteristic length L, characteristic velocity V_∞ in an unsteady motion of characteristic time t_∞. We assume that generally $a \sim O(a_\infty)$, a_∞ being the characteristic sound speed. Then in Equation (VI.10′) appear the dimensionless parameters

$$\text{Mach no. } M_\infty \equiv V_\infty/a_\infty$$

and

$$\text{Strouhal no. } \kappa \equiv L/V_\infty t_\infty.$$

If $\kappa \sim O(1)$, as $M_\infty^2 \to 0$ the equation reduces to the Laplace equation

$$\text{(VI.12)} \qquad \nabla^2\phi = 0.$$

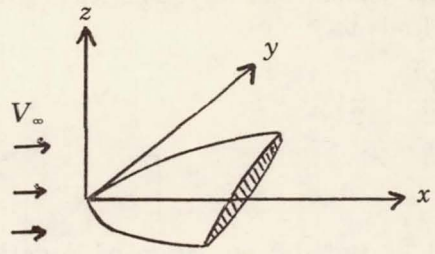


FIGURE 8. Moving Thin Body.

Equation (VI.12) constitutes the (inviscid) "incompressible approximation" since the equation can also be directly obtained by setting $D\rho/Dt = 0$ in the equation of continuity and then using $\mathbf{V} = \nabla\phi$. The velocity potential now may be solved from the prescribed normal derivative of ϕ on the body surface (to satisfy the condition of "no penetration"). After ϕ is obtained, Bernoulli's equation, being an integral of the momentum equation, determines the pressure field which simultaneously must co-exist. We shall not discuss the various techniques of solving Laplace's equation.

VII. **Small perturbation theory for the steady flow over thin bodies.** To illustrate the behavior of the solution of Equation (VI.10′) when the incompressible approximation is not applicable, consider a uniform stream of velocity V_∞ in the x-direction flowing over a fixed thin body lying close to the x, y-plane (see Figure 8). For sufficiently thin bodies, the uniform stream will only be slightly disturbed, and we put the resultant velocity potential as the superposition of that for the uniform stream and a small perturbation, i.e.,

$$\phi = V_\infty x + \phi', \quad \phi'_x, \phi'_y, \phi'_z \ll V_\infty.$$

For the steady case Equation (VI.11) further becomes

$$a^2 - a_\infty^2 = \frac{\gamma - 1}{2}(V_\infty^2 - V^2). \tag{VII.1}$$

After substituting the above into Equation (VI.10′) and retaining only linear terms in ϕ', we get

$$(1 - M_\infty^2)\phi'_{xx} + \phi'_{yy} + \phi'_{zz} \cong 0 \tag{VII.2}$$

provided $|1 - M_\infty^2| \sim O(1)$. By a simple stretching of the coordinates

$$x' = x/\sqrt{|1 - M_\infty^2|}, \quad y' = y, \quad z' = z,$$

Equation (VII.2) reduces to

$$\pm \phi'_{x'x'} + \phi'_{y'y'} + \phi'_{z'z'} = 0,$$

the "$\pm$" corresponds to $M_\infty^2 \lesseqgtr 1$. Thus the subsonic flows all satisfy Laplace's equation while the supersonic flows satisfy the wave equation. In fact, by examining the transformed boundary conditions in the new coordinates, it follows readily that flows over a class of bodies at different Mach numbers can be related to each other. The interpretation of a known flow over a given body and Mach number as that for a different body at a different Mach number is referred to as the "similarity rule". In subsonic flows, such is known as the "Prandtl-Glauert rule", in supersonic flows, the "Ackeret rule".

The linearized theory, Equation (VII.2), fails when some of the neglected terms become comparable with those retained. If we evaluate the neglected terms, it may be verified that the above linearization implies

(i) $a^2 - (V_\infty + \phi'_x)^2 \sim a_\infty^2 - V_\infty^2 \gg V_\infty \phi'_y$ or $V_\infty \phi'_z$

(ii) $a \sim a_\infty \gg \phi'_y, \quad \phi'_z.$

The condition (i) breaks down where $a_\infty^2 \cong V_\infty^2$, or $M_\infty^2 \cong 1$, i.e., in "transonic flows". The condition (ii) breaks down when $a_\infty \ll V_\infty$, or $M_\infty \gg 1$, i.e., in "hypersonic flows". In both transonic and hypersonic cases, then, we are forced to nonlinear theories even for small perturbations.

Let us demonstrate briefly the complications of the transonic approximation. If $V_\infty \cong a_\infty$, it is convenient to consider the flow over a thin body as small perturbations on a uniform sonic flow ($V_\infty = a_\infty = a^*$, say) without the body. Thus putting

$$\phi = a^* x + \phi'$$

and rewriting Equation(VII.1)

$$a^2 = \frac{\gamma + 1}{2} a^{*2} - \frac{\gamma - 1}{2} V^2 \cong a^{*2} - \frac{\gamma - 1}{2} (2a^* \ \phi'_x) + \cdots,$$

we get from Equation (VI.10′) after retaining *all* quadratic terms,

$$-(\gamma + 1)\phi'_x \phi'_{xx} + a^*(\phi'_{yy} + \phi'_{zz}) - 2(\phi'_y \phi'_{xy} + \phi'_z \phi'_{xz}) = 0.$$

The essential features remain unchanged by restricting ourselves to two-dimensional motion in the x, z-plane:

$$-(\gamma+1)\phi'_x\phi'_{xx}+a^*\phi'_{zz}-2\phi'_z\phi'_{xz}=0.$$

Here one or both of the quadratic terms must be everywhere of the same order as the term $a^*\phi'_{zz}$. In order to do so, clearly the function ϕ' must vary much more rapidly in the x-direction than in the z-direction. Hence the first term should dominate, and the "transonic equation" for two-dimensional steady flow finally reduces to

$$-(\gamma+1)\phi'_x\phi'_{xx}+a^*\phi'_{zz}=0. \tag{VII.3}$$

The "similar rule" for relating the transonic flows over geometrically similar bodies of different thickness to each other was deduced by von Kármán.

In analogous manner, the nonlinear perturbation equation for the velocity potential and the similarity rule in hypersonic flow have been given by Tsien. However, strong curved shocks inevitably occur in hypersonic flow, and the flow behind the shock and over the body is generally rotational. Tsien's equation therefore loses much of its significance. On the other hand, if we plot Equation (VII.1) (see Figure 9), in hypersonic flow the sound speed a and the resultant velocity V will always be in the region near the maximum velocity $V_{\max}$

$$V_{\max}=V_\infty\left[1+\frac{2}{(\gamma-1)M_\infty^2}\right].$$

For considerable variation of the local Mach number, the resultant velocity V is essentially unchanged. In addition, the streamlines

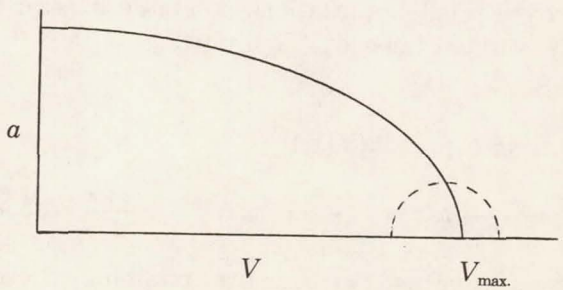


FIGURE 9. Plot of Eq. (VIII. 1).

around thin bodies are always only slightly inclined. Consequently, it is obvious that the perturbation velocity $u' \ll v'$, w'. Neglecting u' completely, the steady flow pattern in the y,z-plane at different streamwise stations x can be interpreted as the unsteady flow pattern in the y,z-plane at successive times, the elapsed time Δt between two stations Δx apart being given by $\Delta t \cong \Delta x / V_\infty$. This is the essence of Hayes' "equivalence principle", which holds regardless of whether the flow is rotational of irrotational, or whether any shock wave occurs at the nose of the body. It simplifies the problem of hypersonic steady flow over a thin body by reducing it to an unsteady flow over a body of lesser dimension.

VIII. **One-dimensional unsteady flow and the formation of shock.** We now return to Equation (VI.10′) but restrict ourselves to one-dimensional unsteady flows,

$$\phi_{tt} - (a^2 - \phi_x^2)\phi_{xx} + 2\phi_x \phi_{xt} = 0. \tag{VIII.1}$$

According to the theory of quasi-linear partial differential equations, this equation is hyperbolic, just as in the acoustic approximation, since the discriminant

$$(2\phi_x)^2 + 4(a^2 - \phi_x^2) = 4a^2 > 0.$$

Thus there exist real characteristic curves, along which the values of ϕ_x and ϕ_t may be described without uniquely determining the higher derivatives ϕ_{xx}, ϕ_{xt}, ϕ_{tt}. Let the running variable along such a characteristic curve be σ. For prescribed ϕ_x and ϕ_t along the curve, the following must hold

$$\begin{aligned} \phi_{x\sigma} &= \phi_{xx} x_\sigma + \sigma_{xt} t_\sigma , \\ \phi_{t\sigma} &= \phi_{tx} x_\sigma + \phi_{tt} t_\sigma . \end{aligned} \tag{VIII.2}$$

We normally should be able to solve ϕ_{xx}, ϕ_{xt} and ϕ_{tt} from Equations (VIII.1) and (VIII.2), except when x_σ t_σ are such that the matrix

$$\begin{pmatrix} 1 & 2\phi_x & -(a^2 - \phi_x^2) & 0 \\ t_\sigma & x_\sigma & 0 & \phi_{t\sigma} \\ 0 & t_\sigma & x_\sigma & \phi_{x\sigma} \end{pmatrix}$$

has rank 2. Hence, to require that the curve be a characteristic

$$\begin{vmatrix} 1 & 2\phi_x & -(a^2-\phi_x^2) \\ t_\sigma & x_\sigma & 0 \\ 0 & t_\sigma & x_\sigma \end{vmatrix} = 0$$

or

$$\frac{dx}{dt} \equiv \frac{x_\sigma}{t_\sigma} = \phi_x \pm a, \tag{VIII.3}$$

giving the direction of the characteristics. Also

$$\begin{vmatrix} 1 & -(a^2-\phi_x^2) & 0 \\ t_\sigma & 0 & \phi_{t\sigma} \\ 0 & x_\sigma & \phi_{x\sigma} \end{vmatrix} = 0$$

or

$$\phi_{t\sigma} = \phi_{x\sigma}(-\phi_x \pm a), \tag{VIII.4}$$

giving a condition on the variation of ϕ_x and ϕ_t along the characteristics. Thus we have two families of characteristics, which may be referred to as the $C_\pm$ curves, respectively, according to the sign "$\pm$" in Equations (VIII.3) and (VIII.4).

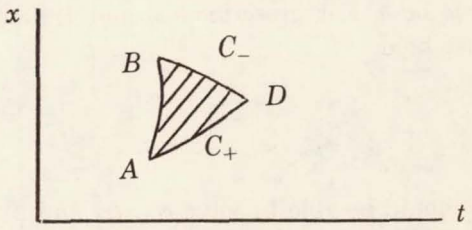


FIGURE 10. Domain of Dependence.

By using the Bernoulli equation, Equation (VI.9), the sound speed "a" may be related to ϕ_x and ϕ_t. If initial data are prescribed along an ordinary curve (not coincident with either characteristic) in the x, t-plane, it is known that the characteristics relations Equations (VIII.3) and (VIII.4) uniquely determine the solution in the curvilinear triangle ABD, bounded by the characteristics C_+ and C_-

through A and B, respectively (See Figure 10). The segment AB is the "domain of dependence" for point D. Likewise, if data are modified along a segment $A'B'$, the solution in the shaded region shown in Figure 11 will be affected and is the "range of influence" of the segment $A'B'$. Moreover, the higher order derivatives normal to a characteristic may be discontinuous. Consequently, a characteristic, and a characteristic only, can serve as the boundary between regions of constant state and variable flow, provided Equation (VIII.1) holds everywhere.

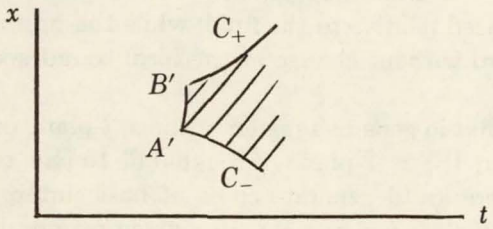


FIGURE 11. Range of Influence.

Now, by differentiating the Bernoulli's equation along a characteristic, there follows

$$\phi_{t\sigma} + \phi_x \phi_{x\sigma} + \frac{2a}{\gamma - 1} a_\sigma = 0.$$

Because of Equation (VIII.4), this reduces to

$$\frac{d}{d\sigma}\left[\pm \phi_x + \frac{2}{\gamma - 1} a \right] = 0.$$

Thus, if we define the "Riemann invariants" r and s as

$$\text{(VIII.5)} \qquad \begin{aligned} r &= \frac{1}{2}\phi_x + \frac{a}{\gamma - 1}, \\ s &= \frac{1}{2}\phi_x - \frac{a}{\gamma - 1}, \end{aligned}$$

it follows that

$$r = r(\alpha), \quad s = s(\beta)$$

where $\alpha = \text{const}$ along the C_+-curves and $\beta = \text{const}$ along the C_--curves. (The running variable σ becomes β along C_+ and α along C_-.) Equivalently, since by Equation (VIII.3)

$$x_\beta/t_\beta = \phi_x + a, \quad x_\alpha/t_\alpha = \phi_x - a,$$

we have

$$\frac{\partial r}{\partial t} + (a + \phi_x)\frac{\partial r}{\partial x} = 0,$$

(VIII.6)

$$\frac{\partial s}{\partial t} + (-a + \phi_x)\frac{\partial s}{\partial x} = 0.$$

The property r is thus propagated *forward* without change at the local sound speed relative to the fluid, while the property s is propagated *backward* without change at the local sound speed relative to the fluid.

It is clear that in general a region in the x, t-plane may be mapped to a region in the r, s-plane through one-to-one correspondence. However, there are degenerate cases of basic interest. If the flow is in a constant state $r = r_0$, $s = s_0$ in a given region in the r, s-plane, this region will be mapped to only a point in the r, s-plane. There may also be regions in the x, t-plane which map into a line $r = r_0$ (or $s = s_0$) in the r, s-plane. The latter case represents motions referred to as "simple waves".

In "simple waves" since the whole region maps to the line $r = r_0$, say, all the s-characteristics (C_--characteristics) become points along $r = r_0$. Back in the x, t-plane, then, along a C_--characteristic $s = s_1$, say, we have

$$r = r_0, \quad s = s_1,$$

hence ϕ_t and ϕ_x must be constants. The C_--characteristics in the x, t-plane therefore will be straight lines.

Let us now consider the flow which is of constant state in a region of the x, t-plane. This region is mapped to a point (r_0, s_0), say, in the r, s-plane. The boundary between this region of constant state and the adjacent region of variable flow must be a characteristic, say $s = s_0$. Now in the r, s-plane all the s-characteristics must start from (r_0, s_0). The next one $s = s_1$ must be located along the line $r = r_0$, since along the boundary $s = s_0$ the characteristics directions extending into the region of variable flow are still completely specified by $r = r_0$. Thus the adjacent region of variable flow must be mapped

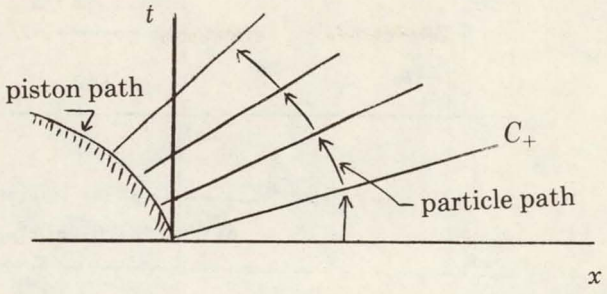


FIGURE 12. Retracting Piston.

into the line segment $(s_0 s_1)^-$ along $r = r_0$. The conclusion is: *The flow adjacent to a region of constant state must be a "simple wave".* "Simple wave" solutions consequently are instrumental in constructing solutions containing regions of constant state.

Consider as example the problem of moving a piston in a long tube filled with gas at rest. The bounding characteristic between the region of gas at rest and the region of moving gas is now a C_+. When the piston is retracting, straight C_+-characteristics can be constructed from the prescribed piston path, as in Figure 12 and the flow completely determined. When the piston is advancing, however, the C_+-characteristics so constructed tend to intersect, as in Figure 13. At the intersection we have different values of r and a given s_0, and the values of ϕ_x and ϕ_t can no longer be solved.

The situation is further clarified by considering a slightly different example. Suppose in a long tube of gas at rest a certain portion is disturbed to the state $r_0(x)$, $s_0(x)$ at $t = 0$. Subsequently, the disturbance r_0 moves to the right at velocity $u + a$, and the disturbance s_0 moves to the left at velocity $u - a$ (see Figure 14). Let the dis-

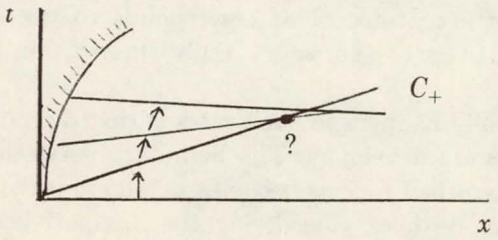


FIGURE 13. Advancing Piston.

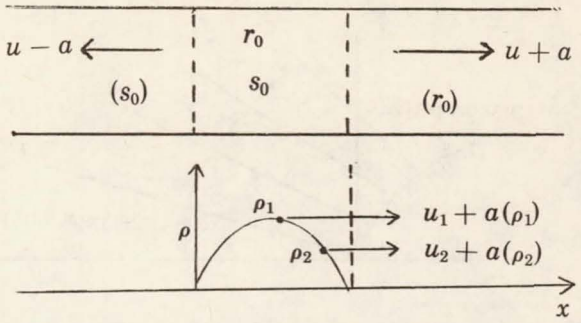


FIGURE 14. Propagation of Initial Disturbance.

turbance r_0 have a density distribution at $t = 0$ as sketched. Since

$$a(\rho) = \sqrt{(\gamma p/\rho)} \propto \rho^{(\gamma-1)/2}$$

we know that $da/d\rho > 0$. Also, as the boundary between the disturbed and undisturbed regions must be a C_+-characteristic, the simple wave solution for the forward propagating disturbance satisfies $s = 0$, i.e., $u/2 - a/(\gamma - 1) = 0$. Hence $du/da > 0$. Thus if $\rho_1 > \rho_2$ as sketched, we conclude: $u_1 > u_2$ and

$$u_1 + a(\rho_1) > u_2 + a(\rho_2).$$

The time history of the density disturbance profile will be as shown in Figure 15, with progressive steepening of the "compression side" of the disturbance (increasing density for the fluid element when swept by the disturbance), and progressive flattening of the "expansion side". Eventually it is seen that the simple wave solution must necessarily break down when the profile develops a vertical slope, since any further progress would require the crest to move ahead of the foot, representing a multi-valuedness of the density which is obviously not acceptable. This corresponds to the situation when the characteristics of the same family intersect in the earlier example.

What actually happens in such cases is that discontinuities in the flow variables are developed. The boundary between the disturbed and the undisturbed regions becomes a "shock wave", instead of a characteristic. Without considering the dissipative mechanisms of the viscosity and heat conductivity of the real gas, the shock wave

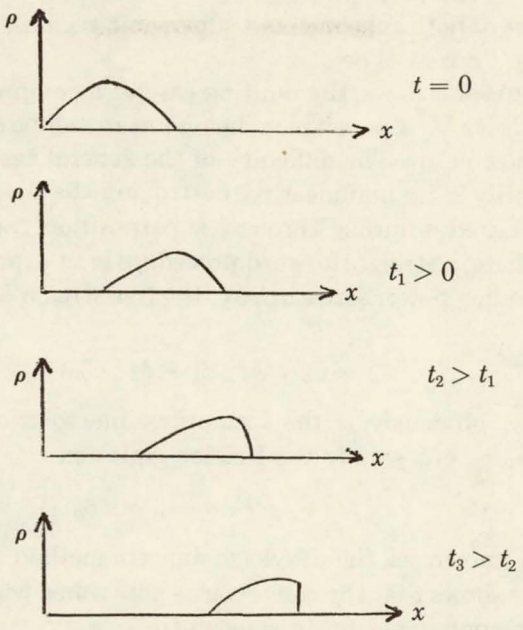


FIGURE 15. History of Density Profile.

is of zero thickness across which finite changes in u and ρ take place. But the basic conservation laws of mass, momentum and energy for the fluid flow in crossing the shock wave must still be obeyed.

IX. **Steady two-dimensional homentropic flows.** If we specialize Equation (VI.10′) to steady two-dimensional flows, the governing equation is

$$(a^2 - \phi_x^2)\phi_{xx} + (a^2 - \phi_y^2)\phi_{yy} - 2\phi_x\phi_y\phi_{xy} = 0. \tag{IX.1}$$

As in the previous section, to classify this equation, the discriminant will be examined. It reads

$$4\phi_x^2\phi_y^2 - 4(a^2 - \phi_x^2)(a^2 - \phi_y^2) = -4a^2[a^2 - (\phi_x^2 + \phi_y^2)].$$

Thus there are three possibilities:

(a) $a^2 - (\phi_x^2 + \phi_y^2) > 0$, i.e., the flow is everywhere subsonic, then the equation is elliptic;

(b) $a^2 - (\phi_x^2 + \phi_y^2) < 0$, i.e., the flow is everywhere supersonic, then the equation is hyperbolic;

(c) $a^2 - (\phi_x^2 + \phi_y^2)$ changes sign in the flow field, which consists therefore of both subsonic and supersonic regions, then the equation is of the "mixed type".

For subsonic flows, the limiting case of incompressible approximation satisfies $\nabla^2 \phi = 0$, which, being linear, can be solved conveniently for most cases. The difficulty of the general case Equation (IX.1) is primarily in its nonlinearity, destroying the possibility of building up a desired solution through superposition. So long as $M^2 < 1$ everywhere, a straightforward procedure is to expand the solution in an ascending power series in, say, the free stream Mach number M_∞, i.e.,

$$\phi = \phi_0 + M_\infty^2 \phi_1 + M_\infty^4 \phi_2 + \cdots$$

where ϕ_0 obviously is the incompressible solution. The successive terms ϕ_1, ϕ_2, etc. satisfy the Poisson equation

$$\nabla^2 \phi_n = F_n(\phi_0, \phi_1, \cdots, \phi_{n-1}).$$

This is known as the Rayleigh-Janzen method. As expected, experience shows that the convergence gets worse when the local Mach number approaches unity somewhere.

More generally, Equation(IX.1) may be reduced to a linear problem by means of a "hodograph transformation", considering x and y as functions of u and v. The continuity equation, Equation(VI.3), may be written as

$$\frac{\partial}{\partial x}(\rho u) + \frac{\partial}{\partial y}(\rho v) = 0$$

from which a stream function ψ may be defined such that

$$\rho_0 \psi_y = \rho u, \quad \rho_0 \psi_x = -\rho v \tag{IX.2}$$

where ρ_0 is a reference constant density. Let now (ϕ, ψ) replace (x, y) as the dependent variables. Also use the variables (q, θ) as polar coordinates for the velocity, i.e.,

$$u = q \cos\theta,$$

$$v = q \sin\theta.$$

Then by definition the following complex relation holds,

$$d\phi + i\frac{\rho_0}{\rho} d\psi = u\,dx + v\,dy + i(-v\,dx + u\,dy)$$

$$= qe^{-i\theta} dz$$

with $z = x + iy$. Hence

$$z_q = q^{-1}e^{i\theta}\left[\phi_q + i\frac{\rho_0}{\rho}\psi_q\right],$$

$$z_\theta = q^{-1}e^{i\theta}\left[\phi_\theta + i\frac{\rho_0}{\rho}\psi_\theta\right].$$

Requiring now that $z_{q\theta} = z_{\theta q}$, we find by equating the real and imaginary parts,

$$\phi_q = q\frac{d}{dq}\left[\frac{\rho_0}{\rho q}\right]\psi_\theta,$$

(IX.3)

$$\phi_\theta = \frac{\rho_0 q}{\rho}\psi_q$$

ρ being here regarded as a function of q. It is now possible to derive a *linear* equation in either ϕ or ψ be elimination. For instance, in terms of ψ,

$$\frac{\partial}{\partial\theta}\left[q\frac{d}{dq}\left(\frac{\rho_0}{\rho q}\right)\psi_\theta\right] - \frac{\partial}{\partial q}\left[\frac{\rho_0 q}{\rho}\psi_q\right] = 0. \tag{IX.4}$$

This equation was first derived by Chaplygin in 1904 in his investigation on gas jets. The disadvantage here is that the boundary conditions involving a given body become very involved. One usually has to take a solution and then find out the exact body shape for which it is the solution.

The relation $\rho(q)$ implied above of course is given by the Bernoulli equation. Now in the incompressible case, Equation (IX.3) reduces to

$$\phi_q = -\frac{1}{q}\psi_\theta, \quad \phi_\theta = q\psi_q.$$

Chaplygin observed that the general case will assume a similar form *if*

$$q\frac{d}{dq}\left(\frac{\rho_0}{\rho q}\right) = -\frac{\rho}{\rho_0 q}$$

which may be integrated into $q^2 \propto 1 - (\rho_0/\rho)^2$, expressing the required $\rho(q)$. Indeed the Bernoulli equation yields such a form for the hypothetical gas with $\gamma = -1$. Thus by approximating the true isentropic relation $p \propto \rho^\gamma$ by an expression

$$p = a + b\rho^{-1}$$

it becomes possible to relate incompressible and compressible solutions *in the hodographic variables*. The well-known Kármán-Tsien approximation for subsonic flows amounts to a tangent approximation of the isentropic curve p against $1/\rho$, near the reference density ρ_0, chosen as the density at the "stagnation point" where $q = 0$.

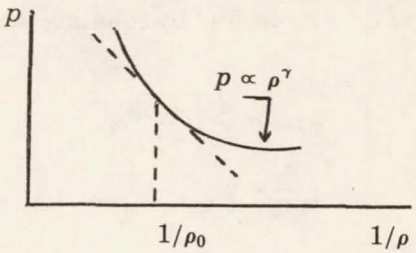


FIGURE 16. Isentropic Curve in Kármán-Tsien Approximation.

It should be noted that Equation (IX.3) is no less general than Equation (IX.1) and therefore not restricted to subsonic flows. As a matter of fact, certain simple solutions gives examples of continuous flows involving both subsonic and supersonic regions. In interchanging the roles of (x, y) and (q, θ), however, a one-to-one correspondence is implied. When $\partial(q, \theta)/\partial(x, y) = 0$ a finite region in the x, y-plane is mapped to a line or a point in the q, θ-plane. We get for the former a solution, if $M > 1$ everywhere, known as the "Prandtl-Meyer flow", corresponding to the "simple wave" of the previous section, while the latter clearly represents a uniform flow. When $\partial(x, y)/\partial(q, \theta) = 0$, a finite region in the hodograph plane is mapped generally to a line in the physical plane, requiring therefore multivaluedness of the flow along this line—again a physically unacceptable situation. Such lines are known as "limit lines", occurring only when $M > 1$ locally and indicating the breakdown of the assumed continuous irrotational homentropic flow.

In the region of supersonic flow, since Equation (IX.1) becomes hyperbolic, the method of characteristics again may be used. The characteristic directions, corresponding to Equation (VIII.3) are found to be

$$\frac{dy}{dx} = \frac{-\frac{uv}{a^2} \mp \sqrt{(M^2-1)}}{1-\frac{u^2}{a^2}} \tag{IX.5}$$

the "$\mp$" sign corresponding to the $C_{\pm}$-directions, respectively in Figure 17. It can be verified that the $C_{\pm}$-directions make an angle θ_M with the local velocity vector, θ_M being the "Mach angle",

$$\theta_M = \sin^{-1}\frac{1}{M}.$$

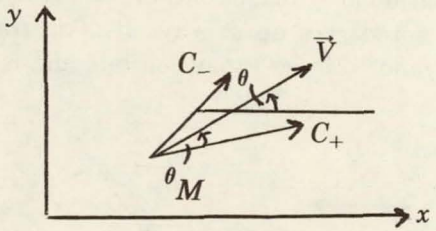


FIGURE 17. Characteristics in Supersonic Flow.

The characteristic conditions, corresponding to Equation (VIII.4), may be conveniently expressed in p and θ, θ denoting the local velocity direction, as

$$\frac{\cot\theta_M}{\rho}\, dp \mp q^2 d\theta = 0 \text{ along } C_{\pm}. \tag{IX.6}$$

The "simple wave" solution when a finite region in the x, y-plane is mapped to a single characteristic C_+, say, in the p, θ-plane follows directly

$$dp = \rho q^2 d\theta / \sqrt{(M^2-1)},$$

hence $dp/d\theta > 0$ in such flows. In conjunction with the Bernoulli equation the above may be integrated. We only note that since

$$\frac{dp}{\rho} + q\, dq = 0$$

the "simple wave" equation may also be written as

$$\frac{dq}{q} = -\frac{d\theta}{\sqrt{(M^2-1)}}$$

hence $dq/d\theta < 0$ in such flows. Thus speed increases as pressure (or density) decreases, and vice versa.

The same argument in the previous section may be followed to prove that a region of uniform supersonic flow can be extended continuously into a region of variable flow only through the simple wave solution. As an example, the supersonic flow turning around a corner is obtained by drawing successive C_--lines from the corner until the velocity leaving the last C_--line has turned through the full angle $\Delta\theta$. Since θ is continuously decreasing in the stream direction, pressure drops and speed goes up as a result. The transition region is known as the Prandtl-Meyer expansion fan, and is shown in Figure 18.

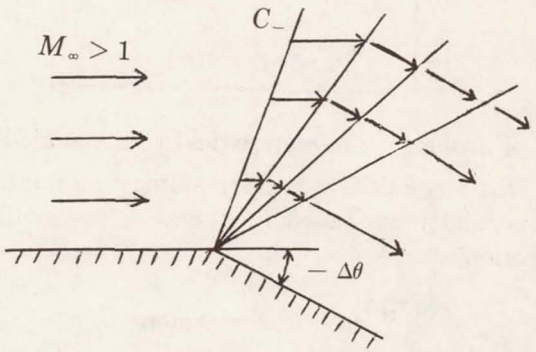


FIGURE 18. Flow Turning a Corner.

If the flow is along a concave wall, $\Delta\theta > 0$ and pressure tends to rise in the streamwise direction. Here again the C_--characteristics will intersect and a continuous solution becomes impossible (See Figure 19). Thus we must expect shock waves to appear in two-dimensional steady flows when a supersonic stream is subjected to a compressive disturbance (increasing pressure in the streamwise direction).

X. **Shock conditions and flows with shocks.** Consider now the one-dimensional problem of a piston moving uniformly at velocity U into a long tube containing gas at rest with pressure p_1 and density ρ_1, as in Figure 20. We postulate a shock wave separating the dis-

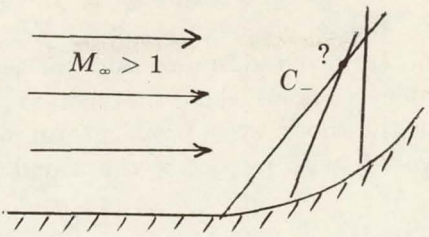


FIGURE 19. Flow Along Concave Wall.

turbed and undisturbed regions, advancing at an unknown velocity U_A. It seems clear that U_A must be defined by only U, p_1, and ρ_1, hence a constant. By dimensional reasoning,

$$U_A = U F\left(\frac{U}{\sqrt{(p_1/\rho_1)}}\right).$$

Now it is possible to let an observer ride on the shock AA, reducing the flow near the shock to a steady one (see Figure 21). For a small area on AA, the conservation laws of mass, momentum and energy then give

$$\rho_1 U_1 = \rho_2 U_2\,,$$

$$p_1 + \rho_1 U_1^2 = p_2 + \rho_2 U_2^2, \tag{X.1}$$

$$\rho_1 U_1\left(h_1 + \frac{U_1^2}{2}\right) = \rho_2 U_2\left(h_2 + \frac{U_2^2}{2}\right),$$

h being the enthalpy as used in Equation (IV.14″). Together with the equation of state, $p = \rho RT$, all the variables with suffix "2" can be solved in terms of those with suffix "1". The result is known as the "Rankine-Hugoniot" relations. It turns out that if $M_1 \equiv U_1/\sqrt{(\gamma RT_1)} > 1$, then p_2/p_1, ρ_2/ρ_1 and T_2/T_1 are all greater than unity, while

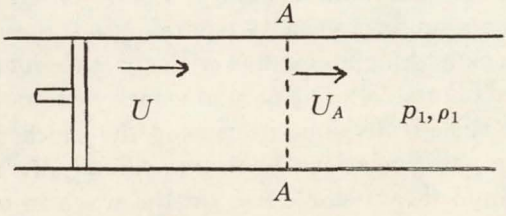


FIGURE 20. Piston Moving in Tube.

$M_2 \equiv U_2/\sqrt{(\gamma RT_2)} < 1$. Furthermore, the entropy S_2 is found to be higher than S_1. Hence, although the same Equation (X.1) holds if both U_1 and U_2 are reversed in direction, the second law of thermodynamics is obeyed only when the motion is in the direction sketched. Thus the shock wave propagates into the calm region at a supersonic speed (with respect to the sound speed ahead of it).

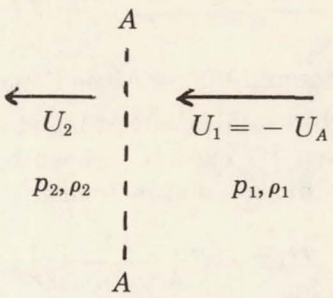


FIGURE 21. Region near Shock Wave.

Turning back to the piston problem, we see that a solution is possible by taking the disturbed region to be in a uniform state, determined by requiring $U_2 = U + U_1$ after the shock of the Mach number M_1. That there is no other solution can also be shown in the following manner: We need a solution in the wedge shaped region in the x, t-plane as shown in Figure 23, taking on the velocity $u = U$ along the line OA and the velocity $U_2 - U_1$ and density ρ_2 from the shock relation along the line OB. Now the lack of a length scale suggests that the solution must depend on a single variable $\xi = x/t$. It is then easily verified that a solution of the form $\rho = \rho(\xi)$, $u = u(\xi)$ cannot be made to satisfy the boundary condition except as constants.

It further is clear that a uniform translation of the entire flow field in any direction should not affect the conservation laws. In particular, by imposing a uniform velocity v parallel to the wave front AA, the oblique shock making an angle α with the oncoming velocity U_1' is obtained (see Figure 23). The normal velocity component is therefore seen to be the effective one in causing the shock wave. In this way the oblique shock relations follow immediately. As the conservation laws are applied to a small area on the wave front in deriving Equation (X.1), the shock relations are actually local in nature and remain valid locally on any curved shock surface.

The real gas of course has viscosity and heat conductivity as

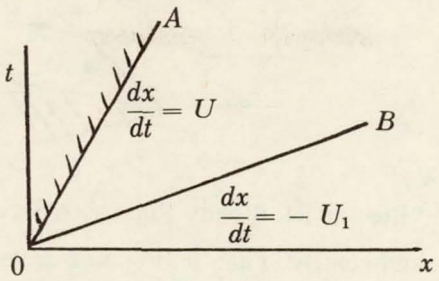


FIGURE 22. Flow Regions in the Piston Problem.

dissipative mechanisms, which resist the discontinuity occurring in a shock wave of zero thickness. The net effect is to smear out the shock wave, so that the upstream and downstream densities ρ_1 and ρ_2, say, are approached only asymptotically. However, most of the change occurs in a very small thickness of the order of the mean free path of the gas molecules. Unless the upstream or downstream part of the flow varies significantly in such a small thickness, the shock structure plays a negligible role in fluid dynamics.

We now mention some examples of steady flows with shock waves. Consider a two-dimensional wedge placed in a uniform supersonic stream. In previous sections, we concluded that the compressive disturbances due to the turning of the streamlines to parallel the wedge surface causes the presence of shock waves. By inserting a straight oblique shock attached to the vertex, it is generally possible to have a uniform and supersonic flow parallel to the wedge after the shock, as shown in Figure 24. For a given $M_1 > 1$, however, the wedge angle may be too large for any attached straight shock to turn

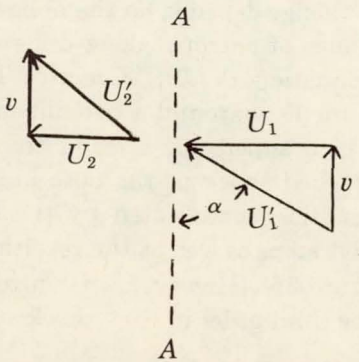


FIGURE 23. Region near Oblique Shock Wave.

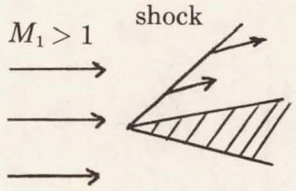


FIGURE 24. Steady Flow with Shock Wave. I.

the streamline sufficiently. Then it becomes necessary to postulate a "detached shock" in front of the body, starting necessarily as a normal shock at the line of symmetry, as shown in Figure 25. The flow behind the normal shock is of course subsonic, but as the shock bends gradually toward the body surface away from the line of symmetry, the flow behind the shock eventually becomes supersonic. The flow problem involving detached shocks is thus of the mixed type and difficult to treat except numerically.

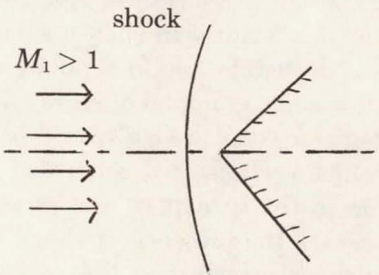


FIGURE 25. Steady Flow with Shock Wave. II.

As soon as curved shocks appear in a uniform stream, it should be noted that the flow behind it is strictly speaking always rotational. The entropy change depends on the obliqueness of the shock, and the different values of entropy along different streamlines give rise to vorticity (Equation (VI.7), "Crocco's Theorem"). For instance, to calculate the flow around a two-dimensional curved body with a sharp nose in a supersonic stream, the solution should be started with an attached shock at the nose and continued by the method of characteristics, complicated by the unknown shock inclination at successive steps as well as the resulting rotational nature of the flow (see Figure 26). However, the entropy change across a shock turns out to be third order in the "shock strength" parameter,

which may be taken as $(p_2 - p_1)/p_1$. If the shock is not strong, the assumption of isentropic flow is not too far wrong. Thus a practical approximate method for the curved body problem, known as the "shock expansion method", is to regard the streamline immediately adjacent to the body as following a Prandtl-Meyer expansion after the leading edge shock. Its use of course cannot be extended to hypersonic flows where the shock will always be quite strong.

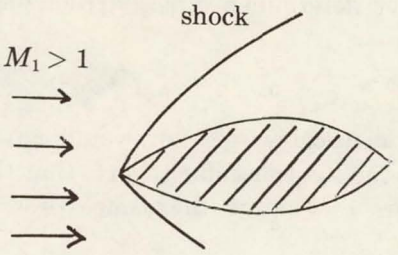


FIGURE 26. Steady Flow with Shock Wave. III.

In hypersonic flow the blunt body is of practical interest (see Figure 27). The very difficult problem of the mixed type flow behind a detached shock is an inherent feature. The limiting case of $M_1 \to \infty$, however, permits at least a much simplified first approximation.

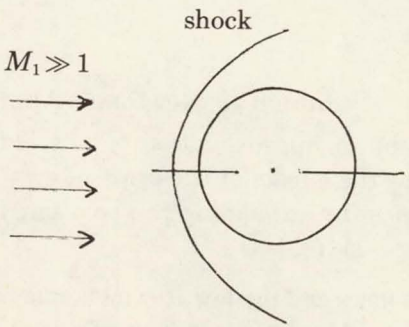


FIGURE 27. Steady Flow with Shock Wave. IV.

Based upon the observation that, after a normal shock the Rankine-Hugoniot conditions give the density ratio as

$$\frac{\rho_2}{\rho_1} = \frac{U_1}{U_2} = \frac{(\gamma+1)M_1^2}{2+(\gamma-1)M_1^2} \to \frac{(\gamma+1)}{(\gamma-1)} \text{ as } M_1^2 \to \infty,$$

it follows that *if* $\gamma = 1$, then $\rho_2/\rho_1 \to \infty$. For air at normal temperatures $\gamma = 1.4$; at very high temperatures with dissociation γ becomes even closer to unity. Thus as a rough approximation we might examine the flow with $\gamma = 1$. Now as $\rho_2/\rho_1 \to \infty$, the shock will locally simply wrap around the body, since the continuity equation is only satisfied by having no thickness between the shock and the body. Neglecting the actual thin "shock layer" thickness, the pressure on the body can be determined directly from simple momentum consideration:

$$p - p_1 \cong \rho_1 U_1^2 \cos^2\theta.$$

This result is identical with what would have been predicted according to Newton's corpuscular theory, that the oncoming gas consists of particles moving at the same speed U_1 (see Figure 28).

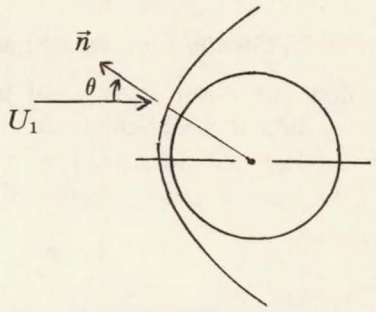


FIGURE 28. Newtonian Approximation.

Hence this type of approximation is referred to as "Newtonian". For refinement the effects of the error of $\gamma = 1$ has been accounted for, for instance, by expanding the solution in terms of the small parameter $(\gamma - 1)/(\gamma + 1)$.

XI. **Viscous flows and the low Reynolds number approximations.** We have so far considered a great deal of the fluid motion under the inviscid approximations, on the basis that for small viscosity and high speeds the viscous effects will be confined to a thin boundary layer immediately adjacent to the body and to a wake behind the body. Such theories obviously can be of no value in connection with the question of skin friction and heat transfer which depend on the details of the motion within the boundary layer. Furthermore, the

precise boundaries of the wake must be known in order to construct the essentially inviscid solution surrounding the body and the wake.

It was first systematically observed by Reynolds in pipe flows that the viscous fluid motion can assume different forms depending essentially on the dimensionless Reynolds number $\mathrm{Re} \equiv VL/\nu$, where V is the characteristic velocity (e.g., mean flow velocity through the pipe), L a characteristic length (e.g., the pipe diameter) and ν the kinematic viscosity of the fluid. On the one extreme, at sufficiently low Re the fluid may move steadily in layers parallel to the pipe axis; on the other, at sufficiently high Re the motion may become time-dependent, irregular and random, but with well defined time averages. The former is referred to as "laminar" motion while the latter is referred to as "fully turbulent" motion. Naturally there is also a "transition" region in which the laminar motion develops into the fully turbulent one. Similar types of motion prevail also in boundary layers. The details of the flow clearly cannot be investigated without first knowing whether the motion is lamina or turbulent. The question of the transition from laminar to turbulent motion is thus of prime importance.

Generally speaking, when the flow is ostensibly governed by physical parameters which are invariant with time, the laminar motion corresponds to the solution of the equations of motion under the assumption of steady flow. If deviations from the steady flow can occur without changing any of the governing physical parameters, the question must be one of stability. As usual the stability problem may be formulated by studying the behavior of perturbations. Unfortunately, mathematically the stability theory is rather difficult even for very simple laminar flows under infinitesimal disturbances. The results are further valid only for the initial breakdown of the laminar flow. Nevertheless the stability theory does provide qualitative correlations between transition and the various physical parameters. The actual beginning of the fully turbulent region however is yet beyond the capability of theoretical prediction. The situation is complicated in addition by the fact that, for bodies in flight, irregularities of the body surfaces and in the free stream all have a profound influence on transition.

The analysis of fully turbulent flow is even more difficult. By putting the instantaneous flow variable as the sum of the "mean" part

plus a fluctuation, equations for the mean motion may be derived from the general equations of motion, but contributions due to the nonlinear interaction of various fluctuations inevitably show up as additional unknowns. For instance, in the mean momentum equation, the momentum transfer due to the fluctuating velocity components through the fixed control surfaces of a fluid element leads to the turbulent or "Reynolds" stresses. In simplified analyses, ad hoc assumptions are made by expressing the Reynolds stresses in terms of other mean flow variables, and a formal solution may then be carried out involving adjustable parameters, which are finally chosen in some way to agree with experimental findings. Such theories are of course semi-empirical in nature, but often unavoidable for practical purposes.

We restrict ourselves in the following to only some of the laminar flow problems. It may be noted that most of the peculiar nature of viscous fluid motion is due to the relative roles of the viscous term and the nonlinear convective terms such as $\mathbf{V} \cdot \nabla \mathbf{V}$ in the equations of motion. Thus not much generality is lost when the complications of compressibility are omitted for brevity. For an incompressible viscous fluid, Equations (III.2) and (III.3), together with Equations (III.9) and (III.11), lead to the following

$$\nabla \cdot \mathbf{V} = 0, \tag{XI.1}$$

$$\rho \left(\frac{\partial \mathbf{V}}{\partial t} + \mathbf{V} \cdot \nabla \mathbf{V} \right) = - \nabla p + \mu \nabla^2 \mathbf{V}$$

known as the "Navier-Stokes equations", in which the viscosity coefficient may be regarded as constant if the temperature range is small. Our purpose is to examine some of its solutions for flows over bodies. The boundary condition on the body is that of "no slip" as discussed in §IV. After the velocity and pressure fields are determined, the temperature field may then be solved separately from the energy equation Equation (III.14), under the "no jump" boundary condition.

Since Equation (XI.1) is nonlinear, an attempt to simplify is naturally that of linearization for small perturbations. Considering therefore an object moving at very low speed in a fluid at rest, we might neglect the quadratic "convective" terms $\mathbf{V} \cdot \nabla \mathbf{V}$ from Equation (XI.1). It follows immediately that

$$\nabla^2 p = 0, \tag{XI.2}$$
$$\nabla^2 \nabla^2 \mathbf{V} = 0.$$

Since the highest order derivatives are not disturbed, it appears that all the boundary conditions for the original equations can be accomodated. This was first used by Stokes to calculate the drag on a sphere moving steadily at a low speed V_∞ in an unbound fluid at rest, and is known as "Stokes' approximation". By using the sphere radius "a" as a characteristic length and the speed V_∞ as the characteristic velocity, an order of magnitude estimate gives

$$\frac{\rho \mathbf{V} \cdot \nabla V}{\mu \nabla^2 \mathbf{V}} \sim \frac{V_\infty a}{\nu} = \mathrm{Re}_a.$$

Thus Stokes' approximation corresponds to the limiting case of $\mathrm{Re}_a \ll 1$, i.e., very low Reynolds numbers based upon the sphere radius "a".

The explicit solution of Stokes' sphere problem, however, leads asymptotically for large r to,

$$\mathbf{V} \sim \mathbf{V}_\infty \left[1 + O\left(\frac{a}{r}\right)\right]$$

where r is the radial distance measured from the center of the sphere, the coordinate axis having been fixed on the sphere. As $r/a \to \infty$, we find in fact

$$\frac{\rho \mathbf{V} \cdot \nabla \mathbf{V}}{\mu \nabla^2 \mathbf{V}} \sim \frac{V_\infty r}{\nu} \to \infty,$$

showing that Equation (XI.2) cannot help but fail as an approximation of Equation (XI.1) at far enough distances, regardless of the smallness of the Reynolds number. In other words, the convective terms eventually take the upper hand as compared with the viscous terms. The seemingly innocent Stokes' approximation is not uniformly valid. In fact, for two-dimensional problems, it is easy to see that Equation (XI.2) must lead asymptotically for large r to,

$$\mathbf{V} \sim \mathbf{V}_\infty \left[1 + O\left(\log \frac{a}{r}\right)\right].$$

In an unbound fluid the condition of uniform stream at large distances cannot be satisfied at all.

The difficulty of Stokes' approximation is perhaps best understood by observing that there are actually *two* characteristic lengths in viscous flow. Besides the geometrical length "a" we also have a viscous length ν/V_∞. In the near field close to the body, the dimensionless distance of interest is indeed r/a, but in the far field away from the body the flow must be expressible in terms of the dimensionless distance $\nu r/V_\infty$ regardless of the body. In general, therefore, two separate approximations are called for, to be matched somehow in an overlapping region where both might be acceptable. In the terms of Lagerstrom, Kaplan and Van Dyke, it is an example requiring the matching of an "inner" and an "outer" expansion. For the technique of "inner" and "outer" expansion, see [7].

The criticism of Stokes' approximation regarding its behavior at large distances from the body was first made by Oseen. As a remedy, Oseen's proposal was to recognize the far field as a small perturbation of the steady uniform stream, hence

$$\mathbf{V} \cdot \nabla \mathbf{V} \cong \mathbf{V}_\infty \cdot \nabla \mathbf{V}' \tag{XI.3}$$

where $\mathbf{V}' = \mathbf{V} - \mathbf{V}_\infty$, the perturbation velocity. Since the rest of the terms in Equation (XI.1) are linear in $\mathbf{V}$, they may all be written without change in terms of $\mathbf{V}'$. Retaining Equation (XI.3) as a first approximation of the convective terms *everywhere*, we get the "Oseen equation" for steady flows—

$$\begin{gathered} \nabla \cdot \mathbf{V}' = 0, \\ \rho \mathbf{V}_\infty \cdot \nabla \mathbf{V}' = -\nabla p + \mu \nabla^2 \mathbf{V}' \end{gathered} \tag{XI.4}$$

again with the same boundary condition as before but expressed in $\mathbf{V}'$. As an approximation for the far field, evidently the primary effects of the convective terms are represented correctly. As the body is approached, the flow will be characterized by the geometrical length "a" and the convective terms still are much smaller than the viscous terms for $V_\infty a/\nu \ll 1$. It is however of uncertain validity in the region where the two types of terms are comparable. For the sphere problem, the drag coefficient from the Oseen approximation is found to be

$$C_D = \frac{6\pi}{\mathrm{Re}_a}\left[1 + \frac{3}{8}\mathrm{Re}_a + O(\mathrm{Re}_a^2)\right]$$

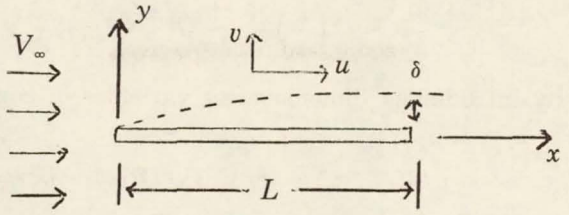


FIGURE 29. Boundary Layer over a Flat Plate.

where the first term agrees with Stokes' result. (Terms up to $O(\mathrm{Re}_a^6)$ have been computed by Goldstein.) A recent more careful analysis shows that the $O(\mathrm{Re}_a^2)$ term in the bracket actually should be $9/40\ \mathrm{Re}_a^2 \log \mathrm{Re}_a$ (see [6]).

XII. **Theory of the boundary layer.** In the other extreme of large Re, we need to describe the motion in the thin "boundary layer" immediately adjacent to the body. In §V it is seen that the boundary layer thickness δ is $O(1/\sqrt{(\mathrm{Re}_L)})$. For a point fixed in space, no matter how close to the body surface, as $\mathrm{Re} \to \infty$ it will lie outside of the boundary layer. This corresponds to dropping formally the viscous terms in the Navier-Stokes equation, and yields the inviscid approximation. In order to keep the point in question within the boundary layer, we must therefore maintain y/δ finite, where y denotes the distance from the body surface, even as $\delta \to 0$ in the limit. To be more specific, consider for simplicity the steady two-dimensional motion of a flat plate moving parallel to itself in an unbounded incompressible fluid. We have here again two length parameters: the geometric length L of the plate, and the viscous boundary layer thickness δ, $\delta \sim L/\sqrt{(\mathrm{Re}_L)}$. In the limit $\mathrm{Re}_L \to \infty$ or $\delta \to 0$, the inviscid solution is simply the undisturbed uniform flow. The u-component velocity in the boundary layer parallel to the plate is generally characterized by V_∞. The order of magnitude of the v-component velocity may be inferred from the continuity equation,

$$v \sim O\left(\int_0^\delta \frac{\partial u}{\partial x}\, dy\right) \sim O\left(\delta \frac{\partial u}{\partial x}\right).$$

Since all changes in the y-direction must be accomplished within the thickness δ, there follows also

$$\frac{\partial}{\partial y} \sim O\left(\frac{L}{\delta}\right) \frac{\partial}{\partial x} \sim O(\sqrt{(\mathrm{Re}_L)}) \frac{\partial}{\partial x}.$$

Thus by introducing dimensionless variables of comparable magnitudes

$$u^* = u/V_\infty, \quad v^* = (v/V_\infty)\sqrt{(\mathrm{Re}_L)}, \quad p^* = p/\rho V_\infty^2,$$
$$x^* = x/L, \quad y^* = (y/L)\sqrt{(\mathrm{Re}_L)},$$

Equation (XI.1) becomes

$$\frac{\partial u^*}{\partial x^*} + \frac{\partial v^*}{\partial y^*} = 0,$$

$$\text{(XII.1)} \qquad u^* \frac{\partial u^*}{\partial x^*} + v^* \frac{\partial u^*}{\partial y^*} = -\frac{\partial p^*}{\partial x^*} + \frac{\partial^2 u^*}{\partial y^{*2}} + O\left(\frac{1}{\mathrm{Re}_L}\right),$$

$$O\left(\frac{1}{\mathrm{Re}_L}\right) = -\frac{\partial p^*}{\partial y^*}.$$

We now let $\mathrm{Re}_L \to \infty$ and omit the terms $O(1/\mathrm{Re}_L)$. The result is the boundary layer equation of Prandtl. The most important feature is the replacement of $\nabla^2 u$ by $\partial^2 u/\partial y^2$ in the x-momentum equation, retaining at least one of the highest order derivatives in the full equation. The solution is also greatly simplified by the consequence of the approximate y-momentum equation that leads to

$$p^* = p^*(x^*),$$

i.e., constant pressure across the boundary layer at each streamwise station. In the flat plate case under consideration the pressure must agree with that in the free stream, hence a constant. The derivation, however, obviously is not valid wherever u and v are of the same order of magnitude, such as the stagnation point regions at the leading and trailing edges of the flat plate.

The same order of magnitude arguments can be applied to bodies of arbitrary but smooth shape. By interpreting the x-coordinate as running along the body surface and y normal to it, the same boundary layer equations result except that the omitted terms include those of $O(\kappa\delta)$, where κ is the characteristic curvature of the body shape. The pressure remains unchanged in the y-direction, the correction due to centrifugal force being $O(\kappa\delta)$. To match the boundary

layer solution with the inviscid solution which prevails beyond the boundary layer, it is noted that since $\delta \to 0$ in the limit, the conditions at the "edge of the boundary layer" must agree with the inviscid solution evaluated at the body surface. This consideration leads to the boundary conditions that, as $y^* \to \infty$

$$u \to u_i(x, 0), \quad p \to p_i(x, 0)$$

$u_i(x, y)$ and $p_i(x, y)$ being the inviscid solutions. There is on the other hand no condition on v^* as $y^* \to \infty$; so long as it is finite, the discrepancy between $v = \lim_{\mathrm{Re}_L \to \infty} v^*/\sqrt{(\mathrm{Re}_L)}$ and $v_i(x, 0) = 0$ is of no consequence at this level of approximation. At the body surface, $y = 0$, the "no slip" condition of viscous fluids must be satisfied by setting $u = 0$, $v = 0$ as usual.

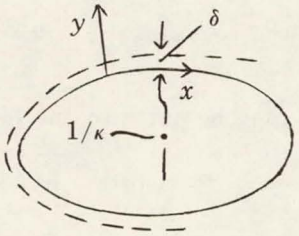


FIGURE 30. Boundary Layer over Curved Body.

It should be remarked that the boundary layer equation would assume different forms depending on the choice of the coordinate system, hence also the flow field which follows as the solution. In his study of the two-dimensional steady incompressible boundary layers, Kaplan introduced the notion of an "optimal" system of coordinates that render the boundary layer solution to agree completely, as $y^* \to \infty$, with the inviscid solution evaluated at the surface $y \to 0$, in both u- and v-components to $O(1/\sqrt{(\mathrm{Re}_L)})$. But the boundary layer solutions in the optimal and any other nonoptimal system of coordinates are shown to be transformable into each other. Furthermore, he proved that the skin friction at the body surface is independent of the coordinate system. The choice of the coordinate system is therefore not too crucial for ordinary purposes.

Though much simplified from the full Navier-Stokes equation, the nonlinear boundary layer equation still defies general treatment. To reduce Equation (XII.1) to a single dependent variable, the

stream function ψ may be introduced by defining

$$u^* = \partial\psi/\partial y^*, \quad v^* = -\partial\psi/\partial x^*$$

guaranteeing thereby the satisfaction of the continuity equation. Now we apply the "von Mises transformation" to the second equation of Equation (XI.1) by choosing (x, ψ) as the independent variables instead of (x, y) and obtain

$$u^* \frac{\partial u^*}{\partial x^*} = -\frac{dp^*}{dx^*} + u^* \frac{\partial}{\partial \psi}\left(u^* \frac{\partial u^*}{\partial \psi}\right), \tag{XII.2}$$

$$p^* = p^*(x), \quad \text{given.}$$

This equation is clearly parabolic in nature. Indeed, if the dimensionless stagnation pressure P^* replaces u^* as the dependent variable

$$P^* \equiv p^* + \frac{u^{*2}}{2}$$

Equation (XII.2) may be put into the form

$$\frac{\partial}{\partial x^*} P^* = [P^* - p^*]^{1/2} \frac{\partial^2 P^*}{\partial \psi^2}$$

which becomes the diffusion equation but with a complicated diffusivity. Thus the solution $P^*(x, \psi)$ can be found in the region $x^* \geqq 0$, $\psi \geqq 0$ if we specify an initial profile $P^*(0, \psi)$ along $x^* = 0$ and also the condition $P^*(x, 0)$ along $\psi = 0$. Since $p^*(x^*)$ is a given function, the conditions on P^* of course are equivalent to the statement that from a given initial velocity profile $u(0, y)$ the solution can be continued uniquely to $x > 0$ for given $p(x)$. As $\psi \to \infty$, $\partial^2 P^*/\partial\psi^2 \to 0$, hence $\partial P^*/\partial x \to 0$ and the solution merges with the inviscid potential flow $P^* = \text{const}$. Note also that no disturbances are propagated upstream.

It is of interest to be able to stipulate an initial profile for the boundary layer over an arbitrary body. For all blunt-nosed bodies there always exists a front stagnation point O. Locally the flow is equivalent to that against an infinite wall normal to the stream. This is recognized as a special case of the symmetrical flow against a wedge of half angle α. (In fact, the flat plate also is a special case, with $\alpha = 0$.) In all such cases it turns out that the inviscid potential flow is of the type

$$V_\infty \propto x^m,$$

the exponent m depending on the half angle α, $0 \leqq m \leqq 1$ for $0 \leqq \alpha \leqq \pi/2$. Although we do not have an initial profile $u(0,y)$, a class of "similar solutions" can be found in the form

$$\frac{u}{V_\infty(x)} = \frac{dF}{d\eta}$$

where $F = F(\eta)$ and η is defined by

$$\eta = y/[2\nu x/(1+m)\,V_\infty]^{1/2}.$$

The function $F(\eta)$ corresponds to the stream function, satisfying an ordinary differential equation:

$$F''' + FF'' + \beta(1 - F'^2) = 0, \qquad \beta = 2m/(1+m) \tag{XII.3}$$

with the required boundary conditions:

$$\begin{aligned} &y = 0, \quad u = v = 0 \text{ or } F(0) = F'(0) = 0, \\ &y \to \infty, \quad u \to V_\infty(x) \text{ or } F'(\infty) \to 1. \end{aligned} \tag{XII.4}$$

The case of $m = 0$ gives the Blasius solution for the flat plate. The case of $m = 1$ gives the stagnation point solution, which is in fact the exact solution of the Navier-Stokes equation for the same problem. The existence and uniqueness of the solution of Equation (XII.3) under boundary conditions Equation (XII.4) were established mathematically for $\beta \geqq 0$ by Weyl. Numerical solutions for various values of β (or m) were calculated by Falkner and Skan, and refined by Hartree. For $\beta < 0$, it is interesting to note that the con-

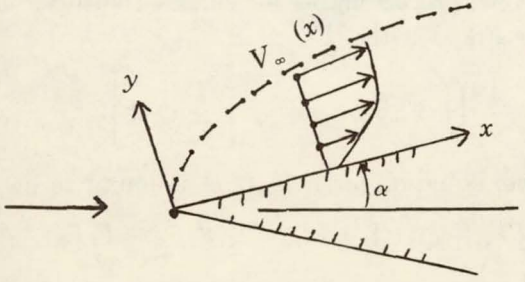


FIGURE 31. Similar Solutions of the Boundary Layer.

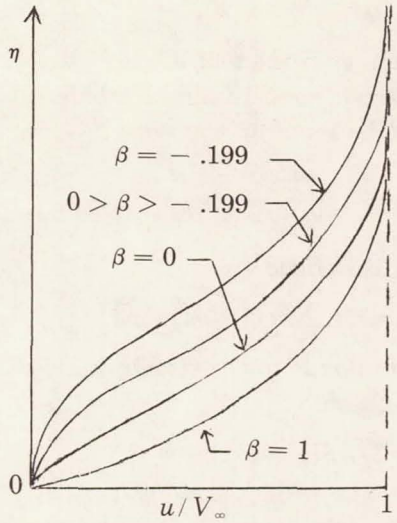


FIGURE 32. Similar Solutions for Various Values of β.

dition $F'(\infty) \to 1$ fails to determine a unique solution, and Hartree had to stipulate the additional requirement that $F'(\eta)$ should approach unity in the most rapid possible way. Even so, he had to stop at $\beta = -.199$, beyond which the velocity develops an overshoot within the boundary layer which seems physically hard to accept. At $\beta = -.199$ the profile has the feature that $\partial u/\partial y|_{y=0} = 0$.

The similar solutions are useful not only to start numerical calculations near the stagnation point, but have often been used as the basis for constructing a first approximation for flows involving rather arbitrary pressure distributions, such as might occur in practical problems. Let an arbitrary $V_\infty(x)$ be given, then $-dp/dx = V_\infty(dV_\infty/dx)$. Now, proceeding as for similar solutions, let $u/V_\infty = \partial F/\partial\eta$, but $F = F(\xi, \eta)$ with

$$\xi \equiv \frac{1}{\nu}\int_0^x V_\infty\, dx, \quad \eta \equiv y\Big/\left[\frac{\nu}{V_\infty}\sqrt{2\xi}\right].$$

The boundary layer equation for F in (ξ, η) is found to be

$$F_{\eta\eta\eta} + FF_{\eta\eta} + \beta(\xi)(1 - F_\eta^2) = 2\xi(F_\eta F_{\xi\eta} - F_\xi F_{\eta\eta}), \tag{XII.5}$$

$$\beta = \frac{2\xi}{V_\infty^2}\frac{dV_\infty}{dx}.$$

The boundary conditions are still Equation (XII.4)

$$\eta = 0, \quad F = F_\eta = 0; \quad \eta \to \infty, \quad F_\eta \to 1.$$

The approximation next is to neglect the right-hand side and solve F as an ordinary differential equation with $\beta(\xi)$ as a parameter. In other words, the similar solution corresponding to the local $\beta(\xi)$ is used as an approximation, the past history being partially accounted for by the ξ-transformation. This is known as the "locally similar" approximation. For improvement Görtler took it as the first term of a series expansion in the solution of Equation (XII.5). Nickel in [5] verified that the local similar solution always provides a lower bound the true solution $u(x,y)$ so long as $d\beta/d\xi \leqq 0$.

Let us adopt the locally similar approximation to get a qualitative picture of the flow within a boundary layer under pressure gradient. If the pressure gradient is always "favorable" ($V'_\infty \geqq 0$), $\beta \geqq 0$, we expect rather normal velocity distributions somewhat like that on a flat plate. But if $V'_\infty < 0$, it becomes possible for β to reach the critical value of $-.199$ and even pass it. There can then be no description of the boundary layer beyond that point. The question arises: what happens then?

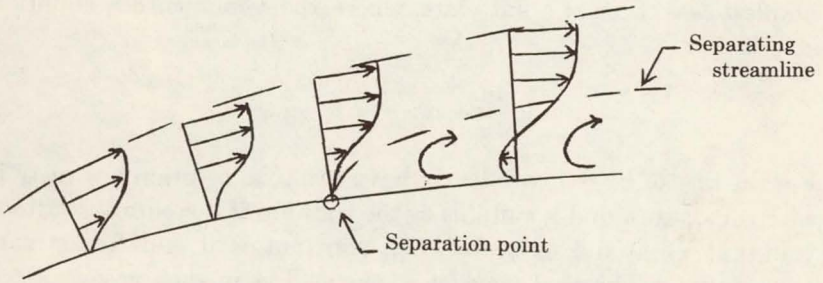


FIGURE 33. Boundary Layer Separation.

It is usually taken that what happens then is the observed phenomenon of "separation", i.e., the streamline begins to detach from the surface. Beyond the separation point, close to the surface the flow direction will be reversed, and the boundary layer approximation ceases to hold. That separation indeed could happen at $\beta = -.199$ is made plausible by noting that here $\partial u/\partial y|_{y=0} = 0$. Since the streamline direction at the wall is given by

$$\frac{dy}{dx} = \frac{v}{u}\bigg|_{y=0} = \frac{\partial v/\partial y}{\partial u/\partial y}\bigg|_{y=0} = -\frac{\partial u/\partial x}{\partial u/\partial y}\bigg|_{y=0}$$

hence $dy/dx = 0$ if $\partial u/\partial y|_{y=0} \neq 0$; but the slope is of the form $0/0$, there indefinite if $\partial u/\partial y|_{y=0} = 0$. According to this criterion, an adverse pressure gradient is *necessary* for separation, in perfect agreement with experience.

We briefly turn again to the energy equation for incompressible fluids, to introduce the basic concepts in heat transfer. The equation reads, after taking the boundary layer approximations,

$$\rho u \frac{\partial h}{\partial x} + \rho v \frac{\partial h}{\partial y} = \frac{\partial}{\partial y}\left(\frac{k}{C_p}\frac{\partial h}{\partial y}\right) = \frac{\mu}{\mathrm{Pr}}\frac{\partial^2 h}{\partial y^2} \tag{XII.6}$$

where the Prandtl number $\mathrm{Pr} \equiv \mu/C_p k \cong .76$ for air in the ordinary temperature range. The equation clearly is of the same structure as the x-momentum equation in Equation (XII.1). When (u, v) are described by a similar solution, h will resemble u in behavior and depend on the same variable. The equation is linear for given (u, v), and can be more easily solved in general. Let us mention only the simplest case, that of a flat plate, where the x-momentum equation is

$$\rho u \frac{\partial u}{\partial x} + \rho v \frac{\partial u}{\partial y} = \mu \frac{\partial^2 u}{\partial y^2}.$$

For the case of $\mathrm{Pr} = 1$, we clearly have a special solution $h = au + b$, with constants a and b, suitable as the solution if h assumes constant values at $y = 0$ and as $y \to \infty$, i.e., constant wall and free stream temperatures. The heat transfer at the wall is, in such cases

$$q_0 \equiv k \frac{\partial T}{\partial y}\bigg|_{y=0} = \mu \frac{\partial h}{\partial y}\bigg|_{y=0} = a\mu \frac{\partial u}{\partial y}\bigg|_{y=0} = a\tau_0$$

or $q_0/\tau_0 = a$, τ_0 being the shear stress at the wall. This is known as "Reynolds' analogy" (of heat transfer and skin friction), ordinarily cast in terms of nondimensional coefficients.

In the case of compressible fluids, the boundary layer concept can be used to derive a set of simplified equations similar to Equation (XII.1). Through the variable density, the momentum and energy equations become coupled and must be solved simultaneously, adding much complexity in the solution. Only in special cases may

the problem be reduced to an equivalent incompressible one through suitable transformations. At hypersonic speeds, the shock wave, whether detached or not, tends to approach the body surface. Then the inviscid and viscous layers would interact with each other, or even merge together. All these phenomena require considerable finesse in handling. For general three-dimensional bodies in unsteady motion, the theory is yet in an undeveloped stage.

References

1. S. Goldstein, *Lectures in fluid dynamics,* Interscience, New York, 1960.

2. H. Jeffrey, *Cartesian tensor,* Cambridge Univ. Press, New York, 1931.

3. H. Lamb, *Hydrodynamics,* 6th ed., Cambridge Univ. Press, New York, 1932.

4. L. M. Milne-Thomson, *Theoretical hydrodynamics,* 4th ed., Macmillan, New York, 1960.

5. K. Nickel, *Eine einfache Abschätzung für Grenzschichten,* Ing. Arch. **31**(1962), 85-100.

6. I. Proudman, and J.R.A. Pearson, *Expansions at small Reynolds numbers for the flow past a sphere and a circular cylinder,* J. Fluid Mech. **2**(1957), 237-262.

7. M. van Dyke, *Perturbation methods in fluid mechanics,* Academic Press, New York, 1964.

CORNELL UNIVERSITY

S. F. Shen

Shock Waves in Rarefied Gases

I. **Introduction.** We shall survey the developments in rarefied gasdynamics toward the solution of flow problems, the shock wave structure serving as an example illustrating the difficulties that led to the various refinements and alternatives. By rarefied gasdynamics is meant the branch of gasdynamics which, because of the effects of very low density, cannot be dealt with by the conventional continuum theory of a viscous and heat-conducting gas, hereafter referred to as simply the (conventional) "continuum theory." The concept of a continuum, however, is usually still adopted, but modifications are required in two main aspects: Firstly, the law relating the viscous stresses to fluid deformation (the Navier-Stokes relations) and that relating the heat flux to temperature gradient (the Fourier law) are theoretically no longer valid. Secondly, the boundary conditions of "no velocity slip" and "no temperature jump" at a solid boundary, generally assumed in conventional continuum theory, must be reexamined. By restricting ourselves to the problem of the shock wave structure, we essentially divorce ourselves from the latter question. Our efforts therefore are directed toward only a formulation of the proper equations to be used in rarified gasdynamics.

To seek a logical theory which is capable of treating the departure from the conventional Navier-Stokes and Fourier laws due to the very low density, we fall back on the kinetic theory of gases. The gas is now regarded as consisting of numerous molecules interacting with each other and with the environment according to the laws of classical mechanics. For the simplest case of a monatomic gas, the molecules are all alike and have spherical symmetry. This will be understood as our model in the following discussion. The phenomenal success of kinetic theory, with suitable chosen force laws between molecules, in predicting quantitatively the viscosity and heat conduction coefficients for use in conventional continuum flows is well known. Equally confirmed are its deductions concerning the flow in the "free molecule" limit of very, very low densities, for instance regarding such flows through orifices or capillaries. These being the two extremes of the spectrum, we expect that it should also be fruitful in intermediate stages that characterize rarefied gasdynamics.

II. **Flow regimes and the Knudsen number.** When we consider a body of gas enclosed in a vessel of volume V, in equilibrium, the state of the gas is defined thermodynamically by the pressure p, the density ρ, and/or the temperature T. These quantities must first be defined from the viewpoint of kinetic theory. Let each of the molecules have a mass m, and the total number in V be N. The density follows directly as

$$\rho = Nm/V = nm \tag{II.1}$$

where $n \equiv N/V$, the number density. If each molecule is characterized by its "size" σ, which may be the diameter for the simplest model of hard sphere molecules, the average spacing λ' of the N molecules occupying volume V is

$$N\lambda'^3 \sim V$$

or $\lambda' \sim n^{-1/3}$. We have thus a dimensionless parameter for the degree of rarefaction of the gas as λ'/σ, i.e., the average size of the cell for each molecule in terms of the molecular diameter. In classical kinetic theory, this parameter is shown to be related directly in the corrections of the perfect gas law:

$$p = \rho RT. \tag{II.2}$$

We shall, however, assume that λ'/σ is sufficiently large that the perfect gas law holds. A typical value of σ is 10^{-8}cm. At standard conditions (0°C and 1 atm.), the number density of gas molecules is given by the Loschmidt number,

$$n = 2.69 \times 10^{19}/\text{cm.}^3$$

or

$$\lambda' \cong \frac{1}{3} \times 10^{-6}\text{cm.}$$

and $\lambda'/\sigma \cong 33$. The ratio becomes obviously larger as the density is decreased, since σ remains unchanged.

The parameter λ'/σ is "static" in nature. As molecules are actually continuously in motion, there is a "dynamic" characteristic length representing the average distance travelled by a molecule between successive collisions, known as the "mean free path" λ. Imagine the molecules arranged actually at distance λ' apart in a regular cubic pattern, as shown in Figure 1. When a molecule moves in an arbitrary direction, the probability of its hitting a second molecule at

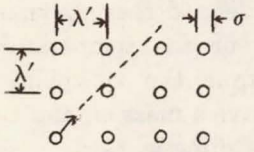


FIGURE 1. The Mean Free Path

a distance of $O(\lambda')$ is proportional to the ratio of the target area σ^2 to the passage area λ'^2. Hence we expect

$$\lambda \sim \lambda' O(\lambda'^2/\sigma^2) \sim O(1/n\sigma^2). \tag{II.3}$$

With the same typical values for λ' and σ the estimate at 0^0C and 1 atm. is

$$\lambda \sim (2.69 \times 10^{19})^{-1} \times 10^{16} \cong \frac{1}{3} \times 10^{-3}\text{cm.}$$

The mean free path goes up quickly as the density is reduced. In the standard atmosphere at 100 miles altitude, for instance,

$$\lambda' \cong 10^{-4}\text{cm.},$$

$$\lambda \cong 10^{4}\text{cm.}$$

Now gasdynamics deals always with flow problems, therefore involving in addition a characteristic length which represents the scale of the flow phenomenon. To fix our ideas, let us imagine the flow as over a body of length L. We assume the flow can be described mathematically as a continuum, so that it must be possible to introduce "fluid elements" of size l with $l \ll L$, as shown in Figure 2. On the other hand, to apply kinetic theory, it is necessary that statistical properties over molecules are well defined in a fluid element. In other words, there must be a sufficient number of molecules in a cell of size l, or $l \gg \lambda'$. Thus rarefied gasdynamics generally deals with the restriction $L \gg l \gg \lambda' \gg \sigma$. Because of the large difference of orders of magnitude between λ and λ', the value of λ may be still taken as arbitrary, compared to l or L.

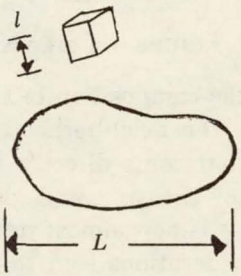


FIGURE 2. Body and Fluid Element

The significance of the mean free path is that it is a measure of the memory of the individual molecules in a flow field of size L. When a molecule hits the body, it rebounds after taking on some characteristics from the body surface. The explicit dependence of the molecular motion on the body characteristics is therefore mainly confined in a "sheath" of thickness roughly $O(\lambda)$ surrounding the body, which is referred to as the "Knudsen layer." Beyond the Knudsen layer the body influence is only indirect, being propagated through successive collisions of the molecules which never were in direct contact with the body. The conventional continuum equations of motion display no explicit dependence on the body geometry and its properties. It seems clear that they are applicable at most to the region beyond the Knudsen layer. If boundary conditions are nevertheless stipulated at the body, the implication must be a vanishing thin Knudsen layer. The case consequently corresponds

to the limit of $\lambda/L \to 0$. The parameter λ/L is known as the Knudsen number, Kn.

In the other limit of $\mathrm{Kn} \to \infty$, the Knudsen layer extends to a sphere of radius λ, the body being shrunk to a small region of size L near the origin, as shown in Figure 3. Of all molecules crossing the spherical surface of area $O(\lambda^2)$, only a small fraction $O(L^2/\lambda^2)$ has collided with the body and rebounded to cross the sphere again.

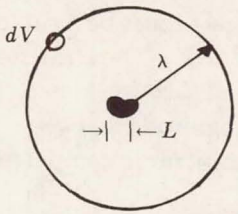


FIGURE 3. Large Kn

Hence, if we examine the composition of the molecules in any small volume element dV in the neighborhood of the spherical surface, there are hardly any that come directly from the body. The flow in the region outside of the sphere is described by conventional continuum theory, but it is now almost undisturbed by the presence of the body. Such considerations lead to the "free molecule flow" approximation, where momentum and energy transfers to the body are evaluated as if the free stream were completely undisturbed by collisions caused by those molecules rebounding from the body. It may be noted, however, that the "free molecule flow" approximation is never valid for the flow field at distances away from the body large compared to the mean free path. In particular, for two-dimensional motions in the x,y-plane, say, many molecules from the z-direction certainly have suffered collisions before arriving at the plane of motion. This has led to difficulties, for instance, in the free molecule flow through a two-dimensional channel.

Between the limits of continuum and free-molecule flows, the flow regimes are often classified according to the magnitude of the Knudsen number, following Tsien [**20**]. Thus for $\mathrm{Kn} \ll 1$ the flow is said to be in the "slip flow" regime, in the sense that only the "no slip" and "no jump" conditions at a solid boundary need to be modified, but the equations of motion remain unchanged. Beyond

the "slip flow" regime and before reaching the free molecule flow limit, i.e., for $\mathrm{Kn} \sim O(1)$, the flow is said to go through a "transition regime." It is in fact for flow problems in this regime that much work remains to be done.

III. **Kinetic definition of pressure and temperature of gas in equilibrium. The Maxwellian distribution function.** We return to the question of defining pressure and temperature for gas in equilibrium in terms of the motion of the molecules. The gas is assumed to be in a fixed vessel, and the pressure is uniform. In the interior consider a small volume element dV enclosed by a control surface S (see Figure 4). Through a small area dS on the surface, molecules

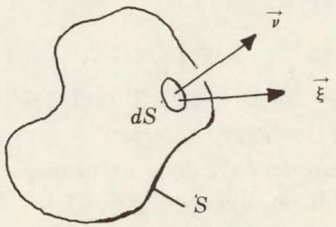


FIGURE 4. Determination of Pressure

having velocity ξ will pass at the rate of $n_\xi \xi \cdot \nu dS$, where n_ξ is the number density of such molecules, and ν the unit outward normal on dS. The rate of momentum loss due to such molecules in the ν direction is therefore $mn_\xi(\xi \cdot \nu)^2 dS$. For all molecules having various velocities ξ, the total rate of momentum loss is obtained by summing over all ξ's. The result on the other hand must be equivalent to the action of a pressure force pdS on the same surface element. Hence we write

$$p = \sum_\xi n_\xi m(\xi \cdot \nu)^2.$$

For a comparable definition of the thermodynamic temperature, we consider again the gas in a fixed vessel of volume V. Let us imagine heat has been added to raise the temperature from absolute zero to T. Kinetically, all this energy, say E, can only go into the translation energy of the monatomic molecules. Thus

$$E = \int_0^T C_v dT \cdot \rho V = \sum_\xi n_\xi V \cdot \frac{1}{2} m\xi^2$$

or

$$\rho \int_0^T C_v dT = \sum_\xi \frac{1}{2} m n_\xi \xi^2$$

where C_v is the specific heat at constant volume, per unit mass, and is generally temperature dependent. This gives an implicit definition of T in terms of the molecular motion.

Since it is always assumed that a large number of molecules are present in the small control volume dV, we must regard ξ as covering the entire range between zero and infinity. The summation over ξ therefore goes over into an integration. The number density n_ξ may be rewritten as

$$n_\xi = nf(\xi)\, d\xi,$$

n being the number density of all molecules, and $f(\xi)\, d\xi$ giving the fraction having velocity between ξ and $\xi + d\xi$. The symbol $d\xi$ above should be understood as a volume element in the velocity space, e.g., in Cartesian space (ξ_1, ξ_2, ξ_3),

$$d\xi = d\xi_1\, d\xi_2\, d\xi_3$$

and not a differential vector. The function $f(\xi)$ is referred to as the "velocity distribution function."

Obviously, since

$$n = \sum_\xi n_\xi = \int_\xi nf(\xi)\, d\xi = n \int_\xi f(\xi)\, d\xi,$$

the velocity distribution function has the property

$$\int_\xi f(\xi)\, d\xi = 1. \tag{III.1}$$

Turning to the pressure definition, we have

$$\begin{aligned} p &= \int_\xi nf\, d\xi m(\xi \cdot \nu)^2 \\ &= \rho \langle \xi_\perp^2 \rangle \end{aligned} \tag{III.2}$$

where $\xi_\perp \equiv \xi \cdot \nu$, the velocity component normal to the surface, and the bracketed quantity $\langle Q \rangle$ represents in general the average of the property Q over all molecules in the velocity space,

$$\langle Q \rangle \equiv \int_{\xi} Q f(\xi)\, d\xi. \tag{III.3}$$

Likewise, for the temperature we get

$$\rho \int_0^T C_v dT = \frac{\rho}{2} \langle \xi^2 \rangle. \tag{III.4}$$

It is next of interest to deduce the velocity distribution function in a gas in equilibrium. First of all there can be obviously no dependence on orientation, so the distribution function must be a function of the speed ξ alone. Let the entire speed range be divided into a finite number of discrete cells according to the average $\xi^{(i)}$ within the cell, and let the number of molecules within the ith cell be a_i. It is then postulated that at equilibrium the distribution of the molecules is the most probable random arrangement of the N molecules into all these cells, subject to the constraints that the total number N and the total kinetic energy E be kept constant.

If we assign sets of numbers a_i, the number of ways to achieve such an arrangement is $N!/\prod_i (a_i!)$. Therefore we write the probability P of such an arrangement as $P = N!/\prod_i (a_i!)$. For convenience, the maximization may be carried out for $P' = \log P$ under the constraints $N = \text{const}$ and $E = \text{const}$. Using Lagrange's multipliers α and α', we finally seek to maximize $P' - \alpha N - \alpha' E$. Since

$$\frac{\partial}{\partial a_i} N = 1,$$

the necessary condition is

$$\frac{d}{da_i} \log(N!) - \frac{d}{da_i} \log(a_i!) - \alpha - \frac{\alpha'}{2} m \xi^{(i)2} = 0.$$

If $x \gg 1$, then with $\Delta x = 1$ we have

$$\frac{d}{dx} \log(x!) \cong \frac{\Delta \log(x!)}{\Delta x} = \log(x!) - \log((x-1)!) = \log x.$$

Hence, assuming $a_i \gg 1$,

$$\log a_i = \log N - \alpha - \frac{1}{2} \alpha' m \xi^{(i)2}$$

or $a_i = NA \exp(-(\alpha'/2) m\xi^{(i)2})$. If the cell size is made to tend to zero formally, the distribution functions must be of the form

$$f = A \exp(-\beta\xi^2) \tag{III.5}$$

known as the Maxwellian distribution function. The two constants A and β are determined by the constant E. To integrate, note that the volume element $d\boldsymbol{\xi}$ should be evaluated as that of the spherical shell between speeds ξ and $\xi + d\xi$, i.e., $4\pi\xi^2 d\xi$. In this way we find

$$A = \left(\frac{\beta}{\pi}\right)^{3/2}, \quad \frac{3}{4\beta} = E = \int_0^T C_v dT. \tag{III.6}$$

Meanwhile, for the pressure p, we choose $\boldsymbol{\nu}$ to be in the Cartesian x-direction and denote the velocity $\boldsymbol{\xi}$ by its Cartesian components (ξ_1, ξ_2, ξ_3). Then

$$p = \rho\langle\xi_1^2\rangle = \rho A \int_{-\infty}^{\infty}\int\int \xi_1^2 \exp(-\beta(\xi_1^2 + \xi_2^2 + \xi_3^2))\, d\xi_1 d\xi_2 d\xi_3 = \frac{\rho}{2\beta}.$$

Since the perfect gas law is assumed to hold, the pressure formula gives

$$\beta = \frac{1}{2RT}. \tag{III.7}$$

From the second of Equation (III.6)

$$C_v = \frac{\partial E}{\partial T} = \frac{\partial}{\partial T}\ \frac{3}{4\beta}\ = \frac{3}{2} R,$$

a well-known result in thermodynamics.

It should be noted that the assumption $a_i \gg 1$ that led to the Maxwellian distribution is clearly violated as $\xi^{(i)} \to \infty$. In fact, although integrations in the velocity space are always formally carried out to $\xi \to \infty$, the logical cut-off for a given E cannot exceed $\xi_{\max} = \sqrt{(2E/m)}$, which is the speed of a single molecule absorbing the entire amount of energy. The assumption of $a_i \gg 1$ ceases to hold before $\xi^{(i)} = \xi_{\max}$, and the Maxwellian distribution function has little significance for molecules whose speeds approach $\xi_{\max}$. It however applies to *almost* all molecules.

With the Maxwellian distribution function, the state of a gas is fully described by the two parameters n and β. Instead of β, it

is often more physical to use the average speed of the molecules $\bar{c}$,

$$\bar{c} \equiv \langle \xi \rangle = \sqrt{\left(\frac{8RT}{\pi}\right)} \tag{III.8}$$

which is quite close to the speed of sound a ($a = \sqrt{(\gamma RT)}$, where $\gamma = 5/3$ for monatomic gases), which is the propagation speed of small disturbances and plays an important role in conventional gasdynamics. Likewise the number density n is directly related to the mean free path λ. The state of the gas molecules thus may also be characterized by $\bar{c}$ and λ. Out of $\bar{c}$ and λ, we can further construct a time constant τ or its inverse Θ:

$$\tau = \frac{\lambda}{\bar{c}} = \frac{1}{\Theta}. \tag{III.9}$$

Θ is known as the "collision frequency" of a molecule, since in time δt, $\Theta \delta t$ gives the average distance travelled $\bar{c}\delta t$ divided by the average distance λ between collisions.

The time constant τ is of considerable interest. If the gas is not in equilibrium, it is plausible to imagine that collisions tend to bring the gas to the most probable, hence the equilibrium distribution. The time constant for this process is no other than τ. In the case of gas in nonuniform motion, there is also a time constant for the overall phenomenon. If the latter is much greater than τ, at each instant and location the molecules in a small volume element dV will be in "quasi-equilibrium." That is, as a first approximation, the velocity distribution should be Maxwellian, with n and β assuming the instantaneous local values, but the observer must now ride with the average velocity U (over all the molecules in dV). We denote this as the "local Maxwellian" $f^{(0)}$,

$$f^{(0)} = \left(\frac{\beta}{\pi}\right)^{3/2} \exp(-\beta[\xi - \mathbf{U}]^2) = \left(\frac{\beta}{\pi}\right)^{3/2} \exp(-\beta \mathbf{c}^2) \tag{III.10}$$

where $\mathbf{c} \equiv \xi - \mathbf{U}$, sometimes referred to as the "thermal velocity." It is easy to verify that following Equation (III.10), $\langle \xi \rangle = \mathbf{U}$ as stated; also $\langle \mathbf{c}^2 \rangle = 3/2\beta$, showing that β (or T) is intimately connected with $\mathbf{c}$.

As a further illustration, in a slightly nonuniform gas let us assume that the state of the molecular motion can be still characterized approximately by $\bar{c}$ and λ. Together with the molecular

properties of mass m and size σ, there are now four parameters from which, among other things, the behavior of the transport properties may be deduced at least qualitatively. For the viscosity coefficient μ, suppose

$$\mu = \mu(m, \sigma, \bar{c}, \lambda).$$

By dimensional reasoning, there must be

$$\mu \sim \frac{m\bar{c}}{\lambda^2} F\left(\frac{\lambda}{\sigma}\right) \sim \rho\lambda\bar{c} G\left(\frac{\lambda}{\sigma}\right).$$

The function $G(\lambda/\sigma)$ should be taken in the limit $\lambda/\sigma \to \infty$. Thus the first approximation should yield $\mu \sim \rho\bar{c}\lambda$, which is confirmed by detailed analysis.

IV. **The Boltzmann equation.** The statement that the velocity distribution function for a gas not in equilibrium is subject to change due to molecular collisions is mathematically expressed by the Boltzmann equation. Consider an arbitrary control volume V enclosed by the surface S in the interior of the gas. On a surface element dS let ν be a unit outward normal (see Figure 5). For

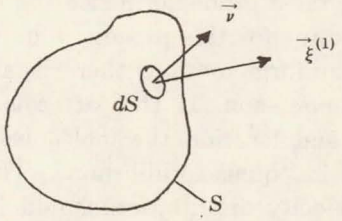


FIGURE 5. Derivation of the Boltzmann Equation

those molecules having velocity between $\xi^{(1)}$ and $\xi^{(1)} + d\xi^{(1)}$, the total number in V at any time is $\int_V nf(\xi^{(1)})\, d\xi^{(1)} dV$. The flux of such molecules through the surface S is $\int_S (\xi^{(1)} \cdot \nu)\, nf(\xi^{(1)})\, d\xi^{(1)} dS$. Let further the rate that such molecules are created in a small volume element dV, through collisions, be denoted by

$$\frac{\partial}{\partial t} nf_1 \bigg|_{\text{coll}} d\xi^{(1)} dV,$$

where $f_1 \equiv f(\xi^{(1)})$. Then we must have

$$\frac{\partial}{\partial t} \int_V nf_1 dV + \int_S nf_1 \xi^{(1)} \cdot \nu\, dS = \int_V \frac{\partial}{\partial t} nf_1 \bigg|_{\text{coll}} dV.$$

By applying Gauss' theorem to the surface integral and letting $V \to 0$, this becomes the Boltzmann equation:

$$\frac{\partial}{\partial t} n f_1 + \nabla \cdot n f_1 \xi^{(1)} = \frac{\partial}{\partial t} n f_1 \Big|_{\text{coll}}. \tag{IV.1}$$

In Cartesian coordinates, since $\xi^{(1)}$ is a constant vector in the derivation, the Boltzmann equation becomes

$$\frac{\partial}{\partial t} n f_1 + \xi_i^{(1)} \frac{\partial}{\partial x_i} n f_1 = \frac{\partial}{\partial t} n f_1 \Big|_{\text{coll}} \tag{IV.1$'$}$$

where x_i are the Cartesian coordinates, $\xi_i^{(1)}$ the Cartesian velocity components of $\xi^{(1)}$, and the convention of summing over identical suffixes is adopted. In more abbreviated form, this is sometimes written as

$$D_1 n f_1 = \frac{\partial}{\partial t} n f_1 \Big|_{\text{coll}}, \qquad D_1 \equiv \frac{\partial}{\partial t} + \xi_i^{(1)} \frac{\partial}{\partial x_i}.$$

The evaluation of the collision terms in the right-hand side of the Boltzmann equation of course requires detailed treatment. A molecular model must be chosen first of all. The dynamic processes of collision are usually simplified by making the following assumptions:

(1) only binary collisions occur, and (2) "molecular chaos" prevails. The second assumption means that the joint probability of finding two molecules having $\xi^{(1)}$ and $\xi^{(2)}$, resp., at a certain place and a certain time is simply the product of the individual probabilities as if the other were absent. Physically, it amounts to the supposition that except during collisions, the molecules are uncorrelated with each other. Both assumptions are valid so long as $\lambda'/\sigma \gg 1$, the second one in particular having been examined in detail by Jeans ([**9**, Chapter IV]).

We shall only briefly sketch what the collision term looks like. The force field of each molecule is taken to be conservative and spherically symmetric. The binary collision between two molecules of velocity $\xi^{(1)}$ and $\xi^{(2)}$ turns $\xi^{(1)}$ into $\xi^{(1)'}$ and $\xi^{(2)}$ into $\xi^{(2)'}$, and may be represented schematically as

$$(1, 2) \to (1', 2').$$

We refer to this as a "direct collision," causing the loss of molecules

"1." Because of the conservative nature of the process, obviously an "inverse collision" can also happen, i.e.,

$$(1', 2') \rightarrow (1, 2),$$

causing a gain of molecules "1." The total number of either type of collisions must depend upon the available number of the participants, as well as the relative speed Ω between the two molecules and "cross section" S representing the effective target area. Thus, the total number of class "1" molecules lost in dV during time δt, through direct collisions with all possible molecules of class "2" is

$$nf_1 dV d\xi^{(1)} \delta t \int_{\xi^{(2)}} nf_2 \Omega S \, d\xi^{(2)} \tag{IV.2}$$

f_2 denoting $f(\xi^{(2)})$. Similarly the gain of molecules of class "1" in dV during δt, through all inverse collisions involving $1'$ and $2'$ is

$$\int_{\xi^{(2)}} nf_1' d\xi^{(1)'} \, dV \delta t \, nf_2' \Omega' S' d\xi^{(2)'}.$$

Note that since $(1, 2) \rightarrow (1', 2')$, for given $\xi^{(1)}$ the inverse collisions must be summed up over all pairs of the $1'$- and $2'$-molecules through the choice of molecules of different $\xi^{(2)}$. However, the details of the collision process show that

$$S' = S, \qquad \Omega' = \Omega, \qquad d\xi^{(2)'} d\xi^{(1)'} = d\xi^{(2)} d\xi^{(1)}.$$

Thus the gain of class "1" through inverse collision can be recast as

$$\int_{\xi^{(2)}} n^2 f_1' f_2' \Omega S \, d\xi^{(1)} d\xi^{(2)} dV \delta t.$$

Consequently, the right-hand side of the Boltzmann equation can be put into a more convenient form and Equation (IV.1) becomes

$$D_1 nf_1 = \int_{\xi^{(2)}} n^2 [f_1' f_2' - f_1 f_2] \Omega S d\xi^{(2)}. \tag{IV.3}$$

In the case of gas in equilibrium, the left-hand side of Equation (IV.3) vanishes. A possible solution is obtained by setting the integrand in the right-hand side to zero. The procedure amounts to the *assumption* that each direct collision is exactly balanced by its inverse, often referred to as the "principle of detailed balancing." As applied to Equation (IV.3), the solution of

$$f_1 f_2 - f_1' f_2' = 0$$

is in general the local Maxwellian distribution $f^{(0)}$, Equation (III.10), and proved to be unique (Jeans [**9**, pp. 25-28]; Grad [**6**]). To satisfy the equilibrium requirement, naturally the mean velocity **U** and temperature T here must be independent of space and time. To justify the "principle of detailed balancing" in this case, it should be mentioned that a consequence of Equation (IV.3) is "Boltzmann's H-theorem," showing essentially that any distribution should indeed tend to $f^{(0)}$ through the collisions. For further discussions of the theorem, see e.g., Chapman and Cowling ([**4**, Chapter 4]).

It is observed that the expression (IV.2) expresses the total number of collisions involving the class "1" molecules. Therefore we may introduce an average collision frequency Θ_1 for a class "1" molecule, and it must be given by

$$\Theta_1 = \int_{\xi^{(2)}} n f_2 \Omega S \, d\xi^{(2)}. \tag{IV.4}$$

Formally then, the Boltzmann equation may also be written as

$$D_1 n f_1 = -\Theta_1 [n f_1 - n \tilde{f}_1] \tag{IV.5}$$

where $\tilde{f}_1$ is to be obtained by identifying the right-hand side with that of Equation (IV.3), and has the significance of an average distribution function for the outcome of the inverse collisions. The property that $\tilde{f}_1 \to f_1^{(0)}$ as the number of collisions increases should be kept in mind.

We note that according to Equation (IV.4), the collision frequency is directly proportional to the number density. In the other limit of $\Theta_1 \to 0$ for extremely rarefied gases the "free molecule flow" is obtained by setting the right-hand side of Equation (IV.5) to zero, neglecting the collisions completely. The property nf then is propagated without change in the direction $\xi^{(1)}$ and at the speed $|\xi^{(1)}|$. By turning Equation (IV.5) into an integral equation, a first order correction for "near free molecule flows" may be obtained through iteration, using the free molecule solution to evaluate $n\tilde{f}_1$, see, e.g., Willis [**21**].

V. **The Maxwell transfer equations and the hydrodynamic equations.** The Boltzmann equation is a nonlinear integro-differential

equation, evidently very difficult to handle. In flow problems, however, the complete information given by the distribution function is much too detailed and more than necessary. Our interest in most cases is in the average properties of all the molecules within a "fluid element," such as the temperature T and velocity $\mathbf{U}$. To deduce equations governing these averages we turn to the Maxwell transfer equations. These are obtained by multiplying the Boltzmann equation by any function $Q(\xi^{(1)})$ and then integrating over the velocity space $\xi^{(1)}$. Since $Q(\xi^{(1)})$ is independent of space and time, the result from Equation (IV.3) is, dropping the superscript (1) on $\xi^{(1)}$,

$$\frac{\partial}{\partial t} n\langle Q\rangle + \frac{\partial}{\partial x_i} n\langle Q\,\xi_i\rangle = \langle \Delta n Q\rangle_{\text{coll}}. \tag{V.1}$$

The term $\langle \Delta n Q\rangle_{\text{coll}}$ on the right-hand side is simply an abbreviation of the rather lengthy expression to be examined immediately.

Written out in full, the term is

$$\langle \Delta n Q\rangle_{\text{coll}} = \int_{\xi^{(1)}} \int_{\xi^{(2)}} Q_1 n^2 [f_1' f_2' - f_1 f_2] \Omega S \, d\xi^{(1)} \, d\xi^{(2)},$$

Q_1 standing for $Q(\xi^{(1)})$. Since the integration is over all $\xi^{(1)}$ and $\xi^{(2)}$, the roles of "1" and "2" may be interchanged without affecting the result. Hence, alternatively,

$$\langle \Delta n Q\rangle_{\text{coll}} = \int_{\xi^{(2)}} \int_{\xi^{(1)}} Q_2 n^2 [f_2' f_1' - f_2 f_1] \Omega S \, d\xi^{(2)} \, d\xi^{(1)}.$$

We next note that when $\xi^{(1)}$ and $\xi^{(2)}$ take on all possible values, so do $\xi^{(1)'}$ and $\xi^{(2)'}$. But,

$$(1, 2) \rightleftarrows (1', 2').$$

If instead all the inverse processes are considered, the integral may also be expressed in

$$\begin{aligned}\langle \Delta n Q\rangle_{\text{coll}} &= \int_{\xi^{(1)'}} \int_{\xi^{(2)'}} Q_1' n^2 [f_1 f_2 - f_1' f_2'] \Omega' S' d\xi^{(1)'} d\xi^{(2)'} \\ &= \int_{\xi^{(1)}} \int_{\xi^{(2)}} Q_1' n^2 [f_1 f_2 - f_1' f_2'] \Omega S \, d\xi^{(1)} d\xi^{(2)}.\end{aligned}$$

Again interchanging the suffixes "1" and "2," we get still another form:

$$\langle \Delta nQ \rangle_{\text{coll}} = \int_{\xi^{(2)}} \int_{\xi^{(1)}} Q_2' n^2 [f_2 f_1 - f_2' f_1'] \Omega S \, d\xi^{(2)} \, d\xi^{(1)}.$$

Finally, a form symmetrical with respect to the indices "1" and "2" is obtained by using the arithmetic mean of the four equivalent expressions:

$$\langle \Delta nQ \rangle_{\text{coll}}$$

$$= \frac{1}{4} \int_{\xi^{(1)}\xi^{(2)}} n^2 [f_1' f_2' - f_1 f_2][Q_1 + Q_2 - (Q_1' + Q_2')] \Omega S d\xi^{(1)} d\xi^{(2)}. \tag{V.2}$$

Equation (V.2) explicitly shows that if Q is a dynamical property which is a "collisional invariant," i.e.,

$$Q_1 + Q_2 = Q_1' + Q_2',$$

then $\langle \Delta nQ \rangle_{\text{coll}}$ vanishes. This is, of course, to be expected. The right-hand side of the Boltzmann equation is the net change of the number of molecules with velocity $\xi^{(1)}$. When multiplied by Q_1 and summed over all the molecules, the result is the net change of the property Q for the aggregate due to collisions. If the sum of Q does not change in any collision, the total cannot change for all the collisions.

For conservative systems the collisional invariants are mass m, momentum $m\xi$ and energy $(1/2)m\xi^2$. With $Q = m$, Equation (V.1) gives the "continuity equation" in conventional gasdynamics,

$$\frac{\partial}{\partial t} \rho + \frac{\partial}{\partial x_i} \rho U_i = 0, \tag{V.3}$$

where U_i are the Cartesian components of the mean or fluid velocity $\mathbf{U}$. With $Q = m\xi$, there follows

$$\frac{\partial}{\partial t} \rho \mathbf{U} + \frac{\partial}{\partial x_i} \rho \langle \xi_i \boldsymbol{\xi} \rangle = 0.$$

Let us write

$$\boldsymbol{\xi} = \mathbf{U} + \mathbf{c}, \text{ or in Cartesian components,}$$
$$\xi_i = U_i + c_i,$$

where $\mathbf{c}$ is the "thermal velocity," and obviously $\langle c_i \rangle = 0$. Then the transfer equation for momentum may be recast in Cartesian coordinates after making use of Equation (V.3), into the conventional form

$$\rho\left[\frac{\partial}{\partial t}U_i + U_i\frac{\partial}{\partial x_j}U_j\right] = \frac{\partial}{\partial x_i}P_{ij} \tag{V.4}$$

where P_{ij} may be identified with the stress tensor in conventional gasdynamics, and is seen to be given from the molecular viewpoint by

$$P_{ij} = -\rho\langle c_i c_j\rangle. \tag{V.5}$$

If we sum the three "normal stresses" P_{11}, P_{22} and P_{33}, Equation (V.5) gives

$$P_{ii} = -\rho\langle c_i^2\rangle = -\rho\langle c^2\rangle.$$

But $\langle c^2\rangle = 3/2\beta = 3RT$, and from the perfect gas law $p = \rho RT$. Hence, as is usually defined, we also have

$$p = -\frac{1}{3}P_{ii}.$$

After taking the pressure out, the kinetic expression for the "viscous stress tensor" is found to be

$$P'_{ij} = -\rho\langle c_i c_j\rangle + \delta_{ij}p \tag{V.6}$$

where $\delta_{ij} = 0$ for $i \neq j$, $\delta_{ij} = 1$ for $i = j$.

Similarly, for $Q = m\xi^2/2$, the transfer equation can be manipulated into the form of the conventional "energy equation,"

$$\rho C_v\left(\frac{\partial T}{\partial t} + U_i\frac{\partial T}{\partial x_i}\right) = -p\frac{\partial U_i}{\partial x_i} + \Phi + \frac{\partial q_i}{\partial x_i} \tag{V.7}$$

where Φ is the "dissipation function,"

$$\Phi = P'_{ij}\frac{\partial U_i}{\partial x_j}$$

and $\mathbf{q}$ is the "heat flux vector," kinetically expressed as

$$\mathbf{q} = -\rho\left\langle\frac{1}{2}c^2\mathbf{c}\right\rangle. \tag{V.8}$$

In general Equations (V.3), (V.4) and (V.7) are called the "hydrodynamic equations," describing the change of the fluid dynamical properties ρ, $\mathbf{U}$ and T in terms of the stress tensor and the heat flux vector. If the distribution function is assumed to be a local

Maxwellian, it is easily verified by direct calculation that $P'_{ij} = q_i = 0$, corresponding to the inviscid and non-heat-conducting approximation. When deviation from the local Maxwellian is small, we shall see that the Navier-Stokes and Fourier laws emerge. When these are no longer sufficient, there seems to be no other course except through further analysis of the Boltzmann equation to achieve an adequate approximation of the distribution function.

VI. **Asymptotic expansion for near Maxwellian distributions.** To begin with, let us recall the Boltzmann equation in the form of Equation (IV.5),

$$D_1 nf_1 = -\Theta_1[nf - n\tilde{f}_1].$$

The left-hand side must have a time constant $\bar{\tau}$ characterizing the change of the function nf_1 as an overall phenomenon. The time constant on the right-hand side is an average of the velocity-dependent collision time $1/\Theta_1$, representing the details. Thus the solution must behave differently depending on the ratio of these two time constants. Let us assume that all Θ_1's are $O(\bar{\Theta})$, an average; the ratio is then $\epsilon = 1/(\bar{\tau}\bar{\Theta})$. When $\epsilon \ll 1$, it is further expected that $f_1 \cong f_1^{(0)}$.

Restricting to $\epsilon \ll 1$, a natural procedure therefore is to seek an asymptotic expansion of f_1 in ascending powers of ϵ:

$$f_1 \sim f_1^{(0)} + \epsilon f_1^{(1)} + \epsilon^2 f_1^{(2)} + \cdots. \tag{VI.1}$$

We now revert to Equation (IV.3), noting that the left-hand side is $O(\epsilon)$, compared to the right-hand side and that $f^{(0)}$ is a solution for $\epsilon = 0$. By substituting Equation (VI.1) into Equation (IV.3), to $O(\epsilon)$ the equation becomes

$$D_1 nf_1^{(0)} = \epsilon \int_{\xi^{(2)}} n^2 [(f_1^{(0)'} f_2^{(1)'} + f_1^{(1)'} f_2^{(0)'}) - (f_1^{(0)} f_2^{(1)} + f_1^{(1)} f_2^{(0)})] \Omega S d\xi^{(2)}.$$

A slightly simpler expression results if we set $\epsilon f^{(1)} = f^{(0)}\phi$. The equation for ϕ is

$$D_1 nf_1^{(0)} = \int_{\xi^{(2)}} n^2 f_1^{(0)} f_2^{(0)} [\phi_1' + \phi_2' - \phi_1 - \phi_2] \Omega S d\xi^{(2)}. \tag{VI.2}$$

The left-hand side being known, Equation (VI.2) is a linear integral equation of the Fredholm type. If $f^{(0)}$ is chosen to be indeed the "local Maxwellian,"

$$\int f^{(0)} d\xi = 1, \qquad \int f^{(0)} \xi \, d\xi = \mathbf{U}, \qquad \int f^{(0)} \mathbf{c}^2 d\xi = \frac{3}{2\beta}.$$

Then ϕ must satisfy the following:

$$\text{(VI.3)} \quad \int f^{(0)} \phi \, d\xi = 0, \qquad \int f^{(0)} \phi \xi \, d\xi = 0, \qquad \int f^{(0)} \phi \mathbf{c}^2 d\xi = 0.$$

Equation (VI.3) turns out to be sufficient to guarantee a unique solution of Equation (VI.2) (Chapman and Cowling [4]). In fact all that is needed eventually is a particular solution satisfying Equation (VI.3). The result is the famous Chapman-Enskog solution.

The particular solution of course depends on the explicit form of the left-hand side of Equation (VI.2). Evaluating $D_1 n f_1^{(0)}$ we find,

$$\text{(VI.4)} \qquad D_1 n f_1^{(0)} = n f_1^{(0)} \left\{ \left(\frac{c^2}{c_m^2} - \frac{5}{2} \right) \mathbf{c} \cdot \nabla \ln T + b_{ij} \frac{\partial U_i}{\partial x_j} \right\}$$

where

$$c_m^2 = \frac{1}{\beta},$$

$$b_{ij} = 2 \frac{c_i c_j}{c_m^2} - \frac{2}{3} \frac{c^2}{c_m^2} \delta_{ij}.$$

This expression enables us to be more specific about the time constant $\bar{\tau}$. From the two terms in the bracket, it is seen that

$$\bar{\tau} \sim O\left(\frac{c_m}{\overline{T}} \frac{\Delta T}{L} \right) \qquad \text{or} \qquad O\left(\frac{\Delta U}{L} \right)$$

where $\overline{T}$ is the characteristic temperature level of the flow, ΔT and ΔU are the temperature and velocity ranges, resp., and L is the characteristic length. In order that $\epsilon = 1/(\bar{\tau}\bar{\Theta}) \cong \lambda/(c_m \bar{\tau}) \ll 1$, we must require $(\lambda/L)(\Delta T/T) \ll 1$ and $(\lambda/L)(\Delta U/c_m) \ll 1$. For a fixed Knudsen number λ/L, the accuracy of the Chapman-Enskog solution improves as $\Delta T/T$ and $\Delta U/c_m$ become smaller.

We shall not go into the details of solving ϕ from Equations (VI.2) and (VI.3), for which the reader should consult Chapman and Cowling [4]. It suffices for our purposes to note that, because of the form of Equation (VI.4), it is possible to represent ϕ by

$$\phi = -A\mathbf{c} \cdot \nabla \ln T - Bb_{ij}\frac{\partial U_i}{\partial x_j},$$

A and B being two scalar functions of the thermal velocity $\mathbf{c}$. Once A and B are determined, by Equations (V.6) and (V.8) the viscous stress tensor and the heat flux vector may thus be calculated. Indeed, these assume the same expressions as the Navier-Stokes and the Fourier laws,

$$P'_{ij} = \mu\left(\frac{\partial U_i}{\partial x_j} + \frac{\partial U_j}{\partial x_i}\right) - \frac{2}{3}\mu\frac{\partial U_i}{\partial x_j},$$

(VI.5)

$$q_i = k\frac{\partial T}{\partial x_i}$$

and the viscosity coefficient μ and the coefficient of thermal conductivity k are obtained from the functions B and A, resp., through quadrature. The Chapman-Enskog solution, i.e., to $O(\epsilon)$ only, brings out thereby the restricted validity of the conventional gasdynamics, as well as providing a theoretical means of evaluating μ and k with suitable choices of the molecular model.

When the solution is carried out to $O(\epsilon^2)$, the details are even more tedious and the resulting formulae for P'_{ij} and q_i are necessarily much more complicated. Because ϕ contains the spatial gradients of temperature and velocity, the left-hand side of the Boltzmann equation now includes second derivatives and products of first derivatives of these mean flow variables. The time derivatives, on the other hand, turn out to be always capable of elimination in favor of the spatial derivatives. The viscous stress tensor and the heat flux vector are thus dependent only on the spatial variations of the mean flow variables. This peculiar set of solutions is sometimes referred to as "normal solutions." When the P'_{ij} and q_i from the solution to $O(\epsilon^2)$ are substituted into the hydrodynamic equations, the result is known as the "Burnett equations." Since higher order derivatives occur, more boundary conditions than for the Navier-Stokes equations are generally required; such conditions

must be formulated for the solution of the problem. But, although to $O(\epsilon^2)$, if Equation (V.1) is regarded as an asymptotic expansion, the Burnett equations do not necessarily provide better accuracy.

We shall now apply the Navier-Stokes equations to the one-dimensional steady shock problem, which may be stated as follows: Given a uniform stream of velocity U_a, pressure p_a, density ρ_a, find the solution which permits a smooth transition into another uniform stream downstream. The final velocity U_b, pressure p_b and density ρ_b must satisfy the conservation laws of mass, momentum and energy, and may thus be regarded as included in the

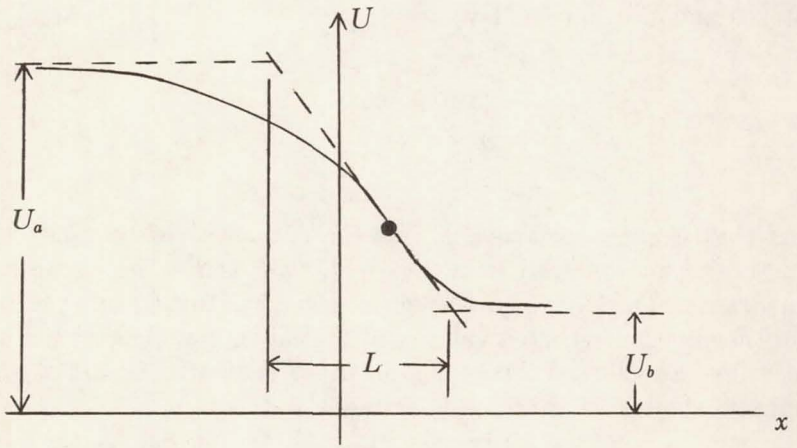


FIGURE 6. Detail of Shock Wave

data U_a, p_a, and ρ_a. As is well known, the relation between U_a and U_b, p_a and p_b, etc. (i.e., between the downstream and upstream quantities) are the Rankine-Hugoniot relations, in which the key parameter is the Mach number,

$$M_a = U_a/\sqrt{(\gamma p_a/\rho_a)};$$

and, to require the entropy change $s_b - s_a > 0$, we must have $M_a > 1$, then $M_b < 1$.

The detailed calculation has been carried out by various authors. Let the attention be focused on a "shock thickness" L defined by the maximum slope and the asymptotic values of one of the

flow variables, say U, as sketched in Figure 6. Without explicitly writing out the governing equations, we are content with some general information from simple dimensional reasoning. The gas is characterized by its material properties μ_a and k_a (both are functions primarily of the temperature for a given gas), and the flow may be characterized by the prescribed U_a, ρ_a, and M_a. But a result of the Chapman-Enskog solution is that μ and k are in fact proportional to each other. Hence the solution must yield

$$L = L(U_a, \rho_a, M_a, \mu_a).$$

By dimensional homogeneity, it follows

$$L = \frac{\mu_a}{\rho_a U_a} F(M_a).$$

Since $\mu_a \sim \rho_a \lambda_a \bar{c}_a$, the above predicts

$$L \sim \lambda_a F(M_a).$$

The detailed calculations give $F(M_a) \sim O(1)$ for all finite $M_a > 1$. Thus we conclude that according to the Navier-Stokes equations the shock thickness is $O(\lambda_a)$—in other words, most of the changes occur within a distance comparable to the (upstream) mean free path. However, *except for weak shocks* where the difference between T_a and T_b or U_a and U_b is small, the restriction $\epsilon \ll 1$ will be violated, and the significance of the result is open to question.

The Burnett equations have been applied by Zoller [**24**] to the shock problem, which happens not to have any ambiguity regarding the boundary conditions. The solution gives a somewhat larger shock thickness at the lower Mach numbers, but predicts oscillations in the profiles at M_a about 1.3 and breaks down when M_a goes beyond about 2. Available experimental evidences do not support the last two rather peculiar results. (See, e.g., Sherman and Talbot [**19**]). As a test case, it is often thought of as an indication that the asymptotic expansion perhaps should not be carried beyond the Chapman-Enskog level.

VII. **Methods based on the moment equations.** Abandoning the expansion in terms of a small parameter ϵ, we look for alternative ways of finding an approximate solution of the Boltzmann equation, Equation (IV.3). Similar to the Rayleigh-Ritz or Galerkin methods

often used in, e.g., vibration problems, a trial function with a number of adjustable parameters may be assumed, and the parameters are to be so chosen that the exact differential equation will be satisfied in some average sense. Now the Maxwell transfer equation, Equation (V.1), may be interpreted as an average of Equation (IV.3) in the entire velocity space when the weighing factor is chosen to be Q, including as special cases the hydrodynamic equations, Equations (V.3), (V.4) and (V.7). The mean flow variables ρ, U_i, and T may be thus interpreted as the "adjustable parameters" present in the trial function, as indeed also the viscous stress tensor P'_{ij} and the heat flux q_i. The hydrodynamic equations unfortunately are too few in number to determine uniquely all these parameters, except when P'_{ij} and q_i are somehow related to the other parameters, for instance, through the Navier-Stokes and Fourier laws. More equations, of course, could be generated by other choices of Q in Equation (V.1). On the other hand, for each Q there seems to be no mathematical reason that the Boltzmann equation must be averaged throughout the entire velocity space. If the velocity space is split in two, say $\xi_1 > 0$ and $\xi_1 < 0$, for every one of Equation (V.1) we get two equations: one from the integration in subspace $\xi_1 > 0$ and one from the integration in subspace $\xi_1 < 0$. All such equations will be referred to as the "moment equations," of which the Maxwell transfer equation itself becomes a special case. By "moment equation method" we mean in general that after assuming a trial function as the approximate solution, the parameters are determined through the choice of a sufficient number of moment equations of one type or another.

For practical reasons the number of parameters in the trial function will have to be rather limited. The accuracy therefore would be quite profoundly affected by the choice of the trial function and the moment equations. It is obviously desirable to incorporate in the trial function as many as possible of the features of the expected solution. There is yet no guidance on how best to choose the moment equations. But, in contrast to the asymptotic expansion, its applicability is not restricted to any special segment of the entire Knudsen number spectrum.

(A) *Grad's thirteen moment equations.* As alluded to in the above, by examining the hydrodynamic equations, it would be

natural to take the flow variables ρ, U_i, T, as well as P'_{ij} and q_i as the parameters in the trial function. This is precisely what Grad proposed to do in [6]. Because of the symmetry the apparent number of parameters is 14. One of these, however, is redundant, since $p = \rho RT = -(1/3)P_{ii}$. The net number of parameters is therefore 13, and eight more moment equations are needed beyond the hydrodynamic equations. Grad's choice was to use again the Maxwell transfer equations but with $Q = c_i c_j$ and $c_i c^2$. Out of the nine equations that result, one of them is also redundant because the "energy equation," Equation (V.7), from $Q = c_i c_i$ is already accounted for.

More specifically, the distribution function which Grad took as the trial function is of the form

$$\text{(VII.1)} \qquad f = f^{(0)}\left[1 + \frac{4}{5}\frac{q_i}{pc_m}\frac{c_i}{c_m}\left(\frac{5}{2} - \frac{c^2}{c_m^2}\right) - \frac{P'_{ij}}{p}\frac{c_i c_j}{c_m^2}\right]$$

where $f^{(0)}$ is the local Maxwellian, and $c_m = 1/\sqrt{\beta}$ as before. It is easily verified that the averages $-\rho\langle c_i c_j\rangle$ and $-\rho\langle(1/2)c_i c^2\rangle$ calculated with Equation (VII.1) are in agreement with the definitions of P'_{ij} and q_i according to Equations (V.6) and (V.8). In fact, it might be remarked that Equation (VII.1) could be written down directly from the Chapman-Enskog solution by replacing the velocity and temperature gradient terms in the latter with P'_{ij} and q_i through Equation (VI.5). Generalization of Equation (VII.1) is possible, as Grad pointed out, by including in the bracket higher order polynomials in c_i orthogonal to the terms present (the "Hermite polynomials"), amounting to a series expansion of the correction to the local Maxwellian in terms of these polynomials. (Additional parameters and moment equations will then be required.) By limiting to Equation (VII.1), the deviation from the local Maxwellian therefore is implied to be relatively small.

With $Q = mc_i c_j$, Equation (V.1) becomes

$$\frac{\partial}{\partial t}\rho\langle c_i c_j\rangle + \frac{\partial}{\partial x_k}\rho\langle c_i c_j \xi_k\rangle = \langle \Delta\rho c_i c_j\rangle_{\text{coll}}$$

or

$$\frac{\partial}{\partial t}\rho\langle c_i c_j\rangle + \frac{\partial}{\partial x_k}\rho U_k\langle c_i c_j\rangle + \frac{\partial}{\partial x_k}\rho\langle c_i c_j c_k\rangle = \langle \Delta\rho c_i c_j\rangle_{\text{coll}}.$$

The left-hand side averages can also be evaluated with Equation (VII.1) but the right-hand side depends on the molecular model. The most convenient model is the so-called Maxwell molecule, which repels another like molecule with a force proportional to r^{-5}, r being the distance between the two molecules. For such molecules, the equation finally may be written as

$$\begin{aligned}&\frac{\partial}{\partial t}P'_{ij}+\frac{\partial}{\partial x_k}U_kP'_{ij}+\frac{2}{5}\left(\frac{\partial q_i}{\partial x_j}+\frac{\partial q_j}{\partial x_i}-\frac{2}{3}\delta_{ij}\frac{\partial q_k}{\partial x_k}\right)\\ &+P'_{ik}\frac{\partial U_j}{\partial x_k}+P'_{jk}\frac{\partial U_i}{\partial x_k}-\frac{2}{3}\delta_{ij}P'_{kl}\frac{\partial U_k}{\partial x_l}\\ &-p\left(\frac{\partial U_i}{\partial x_j}+\frac{\partial U_j}{\partial x_i}-\frac{2}{3}\delta_{ij}\frac{\partial U_k}{\partial x_k}\right)=-\frac{p}{\mu}P'_{ij}\end{aligned}\tag{VII.2}$$

where μ denotes the expression that gives the viscosity coefficient according to the Chapman-Enskog solution for the same molecular model. Likewise, with $Q=(1/2)\,mc_ic^2$, Equation (V.1) leads to

$$\begin{aligned}&\frac{\partial q_i}{\partial t}+\frac{\partial}{\partial x_j}U_jq_i+\frac{7}{5}q_j\frac{\partial U_i}{\partial x_j}+\frac{2}{5}\left(q_j\frac{\partial U_j}{\partial x_i}+q_i\frac{\partial U_j}{\partial x_j}\right)\\ &+RT\frac{\partial P'_{ij}}{\partial x_j}+\frac{7}{2}P'_{ij}\frac{\partial RT}{\partial x_j}-\frac{P'_{ij}}{\rho}\left[\frac{\partial P'_{jk}}{\partial x_k}-\delta_{jk}\frac{\partial p}{\partial x_k}\right]\\ &-\frac{5}{2}p\frac{\partial RT}{\partial x_i}=-\frac{2}{3}\frac{p}{\mu}q_i.\end{aligned}\tag{VII.3}$$

From Equations (VII.2) and (VII.3), we see that in Grad's solution, there is considerable interaction between the stress tensor and the heat flux. More striking when compared with the Chapman-Enskog solution is the presence of the explicit time derivative term in both equations. Thus, if there are no spatial variations, the equations reduce to

$$\frac{\partial}{\partial t}P'_{ij}=-\frac{p}{\mu}P'_{ij},$$

$$\frac{\partial}{\partial t}q_i=-\frac{2}{3}\frac{p}{\mu}q_i.$$

A relaxation phenomenon, nonexistent in the Chapman-Enskog solution, is now predicted. The time constant is $O(\mu/p)$. Since $\mu\sim\rho\bar{c}\lambda$,

$$\frac{\mu}{p} \sim \frac{\rho \bar{c} \lambda}{\rho RT} \sim \frac{\lambda}{\bar{\bar{c}}} \sim \frac{1}{\bar{\bar{\theta}}}$$

which is of course the expected order of magnitude following Equation (IV.5) as discussed in the previous section. But here the result is more quantitative and in particular P'_{ij} and q_i are found to have somewhat different relaxation times.

Of considerable interest is the fact that *both* the Chapman-Enskog and Burnett formulae for P'_{ij} and q_i can be obtained from the Grad equations, even though the Grad equations are obtained from a distribution function, Equation (VII.1), that is at the level of the Chapman-Enskog solution only. This is done by regarding Equations (VII.2) and (VII.3) as definitions for the right-hand side quantities, namely, P'_{ij} and q_i. If P'_{ij} and q_i are small, the effects of the presence of these quantities in the left-hand side may be determined through an iteration process, starting from $P'_{ij} = q_i = 0$. The first iteration then gives precisely the Navier-Stokes and Fourier laws. The second iteration gives the Burnett result, after eliminating $(\partial U_i/\partial t)$ and $(\partial T/\partial t)$ by means of the hydrodynamic equations. This feature of the Grad equations is naturally satisfying. More important to note, however, is that considering the effort in achieving the Chapman-Enskog, not to say the Burnett, solution, it also demonstrates the power of the moment equation method when properly used.

We skip over the question of the boundary conditions for the Grad equations, which have been examined to some extent by Grad himself. As applied to the one-dimensional steady shock problem (Grad [7]) at lower Mach numbers these equations yield solutions which are rather close to the Navier-Stokes result, giving a slightly larger shock thickness; but for Mach numbers greater than about 1.65, again no solution can be found. This is to some extent rather disappointing. The difficulty could only be attributed to the chosen form of Equation (VII.1), which ceases to provide a good approximation when the molecules are far from being in a state of quasi-equilibrium. For lower Mach numbers, i.e., weak shocks, the up- and downstream conditions are not too different from each other, the distribution anywhere within the shock thus deviates little from an average constant Maxwellian. Such is, of course, far from being the case for large Mach numbers and strong shocks.

(B) *Mott-Smith's bimodel distribution.* We now recognize that for strong shocks a trial function not restricted to quasi-equilibrium is necessary. A very simple choice was offered by Mott-Smith in [14], who assumed that it might be taken as a linear combination of the up- and downstream distribution functions,

$$f = \alpha_a(x)\, f_a^{(0)} + \alpha_b(x)\, f_b^{(0)} \tag{VII.4}$$

where $\alpha_a(x)$ and $\alpha_b(x)$ are the adjustable parameters. However, since $\int f\, d\xi = \int f_a^{(0)} d\xi = \int f_b^{(0)} d\xi = 1$, we require

$$\alpha_a + \alpha_b = 1. \tag{VII.5}$$

If the x-axis is in the direction of flow, the boundary conditions are

$$\begin{aligned} &x \to -\infty, \quad && \alpha_a \to 1, && \alpha_b \to 0; \\ &x \to +\infty, && \alpha_a \to 0, && \alpha_b \to 1. \end{aligned} \tag{VII.6}$$

The "bimodal" nature is clear, as for given x the molecules may be regarded as a mixture of two groups maintaining either the up- or downstream characteristics.

There is now in effect only one adjustable parameter. To determine this parameter, Mott-Smith left the hydrodynamic equations alone but employed a moment equation obtained from Equation (V.1) with $Q = \xi_1^2$, or ξ_1^3, which then was solved by imposing Equation (VII.6). The hydrodynamic equations provide as usual the Rankine-Hugoniot relations, expressing all downstream properties in terms of those upstream. A solution for α_a, say, at all $M_a > 1$ was shown to be possible and the flow variables computed as averages. The shock thicknesses so determined from the two choices of Q differ between themselves by 10 to 25 percent depending on M_a. This difference, of course, reflects the uncertainty due to the arbitrariness in choosing Q. There have been consequently discussions attempting to arrive at a criterion for the selection (e.g., Rosen [16], Sakurai [17]). More realistic molecular models have also been used in evaluating the collision terms of the moment equation (Muckenfuss, [15]). At the lower Mach numbers, the Mott-Smith shock thickness is much greater than that from the Navier-Stokes or Grad equations, and is generally considered inaccurate. A comparison is shown in Figure 7, which is taken from [23].

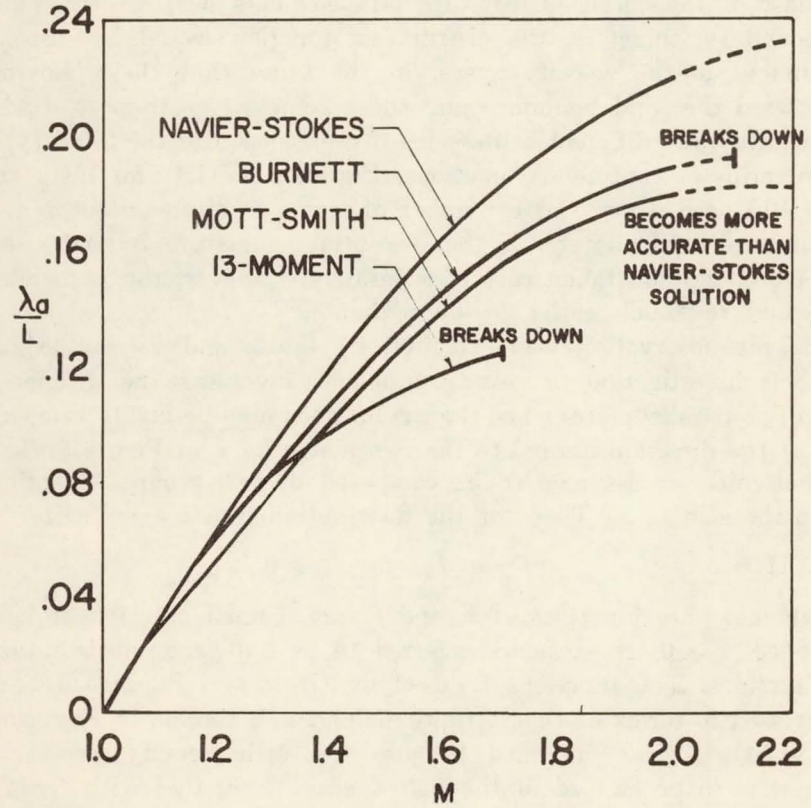


FIGURE 7. Comparison of Reciprocal Shock Thickness for Monatomic Maxwell Molecules[1]

[1] Reprinted from Ziering, S., Ek, F., and Koch, P., Physics of Fluids, 4 (1961), 975-987 with permission of the publishers.

(C) *Methods using half-range distribution function.* We have mentioned that the shock wave is a convenient example in rarefied gasdynamics because of the absence of solid boundaries. When a solid boundary is present, the molecules rebound from, or rather are emitted by the solid boundary, and usually have "forgotten" most of their past history. In fluid elements near to the solid boundary, therefore, the distribution function would be discontinuous in the velocity space, in the sense that those moving toward the solid boundary and those going away from it would require quite different expressions. An expansion of the Grad type in terms of continuous functions, Equation (VII.1) for instance, would need a very large number of terms to approximate a discontinuity adequately. If the discontinuous nature is recognized beforehand and taken care of separately, however, the remainder would be much easier to approximate.

This observation was exploited by Gross and Ziering [8] in their investigation of several problems involving the geometry of two parallel plates when the gas between may be highly rarefied. Let the direction normal to the two plates be x in Figure 8. The molecules are assumed to be composed of two groups according to the sign of ξ_1. Then for the distribution function we write

$$nf = n_{\pm} f_{\pm} \quad \text{for} \quad \xi_1 \gtrless 0 \tag{VII.7}$$

where $n_{\pm}$ are functions of x, and $f_{\pm}$ are defined only in the half spaces $\xi_1 \gtrless 0$, resp., hence referred to as half-range distribution functions. The functions $f_{\pm}$ used by Gross and Ziering are expressed in terms of the Hermite polynomial, similar to Equation (VII.1), slightly modified because the orthogonality condition now is to be applied in the half spaces. Compared with Grad's approach, with the same number of parameters in the expansion,

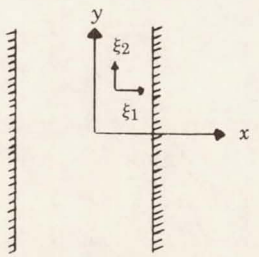


FIGURE 8. Flow between Parallel Plates

evidently the half-range distribution function contains twice as many unknowns; consequently, twice as many moment equations are needed for their determination. Gross and Ziering then split up each of the Maxwell transfer equations, Equation (V.1), into two by carrying out the integration in the two halves of the velocity space separately. In a Grad-like expansion of the half-range distribution function, the adjustable parameters lose the physical significance as corresponding to P'_{ij}, q_i, etc., which are, by definition, the averages over the whole velocity space.

Application of the technique has been limited to several "linearized" problems where the relative velocity or the temperature difference of the two plates is small. In such cases, the half-range distributions were expanded around a constant Maxwellian, and the calculation was rather straightforward.

An alternative choice of the half-range distributions in Equation (VII.7) is the "two-stream Maxwellian" distribution proposed by Lees [11]. The form is taken to be

$$f_{\pm} = \left(\frac{\beta_{\pm}}{\pi}\right)^{3/2} \exp[-\beta_{\pm}(\boldsymbol{\xi} - \mathbf{u}_{\pm})^2] \tag{VII.8}$$

where $\beta_{\pm}$ and $\mathbf{u}_{\pm}$ are the adjustable parameters, in addition to $n_{\pm}$, to be determined by the moment equations. In fact, to generalize the method Lees adopts a "line of sight" principle which divides the molecules into groups as if in free molecule flow. In the problem of the gas between parallel plates, there are thus the same two groups as in the half-range representation of Gross and Ziering, each moving toward one of the walls. For the case of an arbitrary body moving in an unbounded gas region, at the given point P a pencil of rays may be drawn to form a cone tangent to the body, as shown in Figure 9. These molecules in a volume element at P

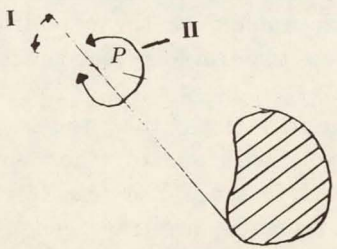


FIGURE 9. The Line of Sight Principle

coming from the body, if in free molecule flow, would have their velocity lying within the cone I. These are taken by Lees as group I. The rest are all taken as group II. The "dual-range" character of the distribution is then expressible in the same form as Equations (VII.7) and (VII.8), except replacing suffices "$\pm$" by suffices "I" and "II." For the needed moment equations Lees prefers to maintain the whole-range transfer equations, Equation (V.1), with successive Q's similar to Grad's that led to the thirteen moment equations. In contrast to Grad's distribution, there are now, however, only ten parameters $n_{\mathrm{I,II}}$, $\beta_{\mathrm{I,II}}$ and $\mathbf{u}_{\mathrm{I,II}}$, none of which, it may be noted, has any significance as physical observables except in dimension. Regarding the transfer equations corresponding to Equations (VII.2) and (VII.3) as expressions for P'_{ij} and q_i in terms of the left-hand side terms, we are inclined to conclude that, together with the hydrodynamic equations, there still should be thirteen equations for the thirteen variables $(\rho, U_i, T, P'_{ij}, q_i)$ at this level of approximation. In general the ten parameters inherent in the "two-stream Maxwellian" appear too few in number to be really self-consistent. When applied to the parallel plates problem with large relative velocity, some difficulty was indeed experienced by Lees and Lin [12]. Their mention of the possible improvement by using skewed "two-stream Maxwellians" amounts to an effort toward additional degrees of freedom.

A further point of criticism may be directed at the "line of sight" principle. The grouping of molecules following this principle is of course correct in the free molecule limit or very close to the body surface, but the principle seems to be rather irrelevant after the molecules have gone through several collisions. Its consequences therefore need not agree with the result from the Navier-Stokes equation in the conventional continuum limit. This drawback is illustrated in the problem of the cylindrical Couette flow (in the annulus between two rotating concentric circular cylinders) investigated by Ai [1].

In spite of these objections, the "two-stream Maxwellian" is relatively easy to work with and together with the "line of sight" principle can be used to set up, at least formally, the governing equations for flows involving arbitrary geometry and large deviations from quasi-equilibrium. It would be of interest to see the

solution of the shock problem by this method, which unfortunately is not available.

VIII. **The BGK model equation and the shock solution.** We have discussed above some of the approximate methods of handling the Boltzmann equation. An entirely different approach is to try for exact solutions by simplifying the Boltzmann equation itself. The most well known of such simplifications is the BGK or Krook model (Bhatnagar et al., [2], Krook [10]). Looking back, we have the Boltzmann equation, Equation (IV.5),

$$D_1 n f_1 = -\Theta_1 [n f_1 - n\tilde{f}_1].$$

The complications are all contained in the right-hand side terms, which will now be approximated.

First of all, the dependence of Θ_1 on the molecular velocity $\xi^{(1)}$ is clearly a matter of detail. It seems reasonable to approximate it with simply $\bar{\Theta}(\mathbf{r}, t)$, an average for all molecules. To simplify the very complicated $\tilde{f}_1$, the choice is made so as to preserve the following important properties of the exact equation:

(a) As $\bar{\Theta} \to \infty$, $f \to f^{(0)}$, the local Maxwellian.

(b) In the transfer equation, Equation (V.1), $\langle \Delta n Q \rangle_{\text{coll}} = 0$ for the collisional invariants $Q = m,\ m\xi,\ mc^2/2$.

(c) There is an "H-theorem."

Krook took directly $\tilde{f} = f^{(0)}$; the model equation is thus

$$D_1 n f_1 = -n\Theta_1 (f_1 - f_1^{(0)}). \tag{VIII.1}$$

That the right hand satisfies the requirements (a) and (b) is immediately obvious. It can be shown that condition (c) is also fulfilled. The rate of change of the function nf is now proportional to its departure from the quasi-equilibrium distribution $f^{(0)}$. Hence Equation (VIII.1) may be regarded as a relaxation model. The equation is, however, only apparently linear, since the parameters β and $\mathbf{U}$ in $f^{(0)}$ remain to be averaged over the unknown f.

All the previous approximate methods of treating the Boltzmann equation can, of course, be applied to Equation (VIII.1). The Chapman-Enskog type of solution, for instance, is obtained by writing

$$f = f^{(0)} + \epsilon f^{(1)} + \cdots.$$

By substitution into Equation (VIII.1), the solution for $\epsilon f^{(1)}$ is explicitly given as, dropping subscript "1,"

$$\epsilon f^{(1)} = \frac{1}{n\bar{\theta}} Dnf^{(0)}. \tag{VIII.2}$$

The dependences on the mean flow gradients $\nabla \ln T$ and $\partial U_i/\partial x_j$ follow from the same term $Dnf^{(0)}$ as in the Chapman-Enskog solution. The Navier-Stokes and Fourier laws are consequently recovered, except that the viscosity coefficient and the coefficient of thermal conductivity are more crudely predicted.

The thirteen moment equations of Grad can also be derived for Equation (VIII.1). The left-hand sides of Equations (VII.2), (VII.3) are unchanged if Equation (VII.1) is maintained. The right-hand sides depend on the details of collisions but with Equation (VIII.1) they can be written down by inspection. The counterparts to Equations (VII.2) and (VII.3) are thus found to be

$$\begin{aligned} \frac{\partial}{\partial t} P'_{ij} + A_{ij} &= -\bar{\theta} P'_{ij}, \\ \frac{\partial}{\partial t} q_i + B_i &= -\bar{\theta} q_i \end{aligned} \tag{VIII.3}$$

where A_{ij} and B_i stand for all the terms except the time derivative in the left-hand sides of the corresponding Grad equations. The only difference lies in replacing the two relaxation times μ/p and $(3/2)(\mu/p)$ with a single time constant $1/\bar{\theta}$. For this reason, the Krook approximation is sometimes referred to as the single relaxation model. The comparison also suggests that the average $\bar{\theta}$ may be taken to be

$$\bar{\theta} = \frac{p}{\mu} \quad \text{or} \quad \frac{2}{3}\frac{p}{\mu} \tag{VIII.4}$$

depending on whether P'_{ij} or q_i is the dominant feature.

In the near continuum regime which is adequately described by the Grad equations, the difference between the BGK model and the Boltzmann equation amounts thus to a difference in the Prandtl number $\mathrm{Pr} \equiv \mu C_p/k$. The correct value is 2/3 for monatomic gases while from the BGK model the Prandtl number will be unity. The regime of free molecule flow in the limit $\theta_1 \to 0$ is un-

affected by the approximation. Its validity in the transition regime is rather difficult to assess, although the common belief is that it should serve as a reasonable interpolation.

The integral equation form of Equation (VIII.1) has been the basis for a number of applications. For brevity consider the steady flow problem:

$$\xi^{(1)} \frac{d}{ds_1} n f_1 = -\bar{\Theta} n [f_1 - f_1^{(0)}] \tag{VIII.5}$$

where s_1 is the distance along the direction of $\xi^{(1)}$. Direct integration yields, after dropping the subscript "1,"

$$nf = n'f' \exp\left[-\int_{s'}^{s} \left(\frac{\bar{\Theta}}{\xi}\right) ds \right] + \exp\left[-\int_{s'}^{s} \left(\frac{\bar{\Theta}}{\xi}\right) ds \right] \int_{s'}^{s} nf^{(0)} \exp\left[\int_{s'}^{s} \left(\frac{\bar{\Theta}}{\xi}\right) ds \right] \left(\frac{\bar{\Theta}}{\xi}\right) ds \tag{VIII.6}$$

where the boundary condition $n'f'$ at $s = s'$ is assumed given. The integral $\int_{s'}^{s} (\bar{\Theta}/\xi)\, ds$ represents the number of collisions for such molecules in traveling the distance between s and s', and the exponential factor is the probability that the molecules from s' should survive. The second term is the gain of such molecules as collision products. Since the unknown functions n and f are involved in the latter, usually an iterative procedure is necessary to achieve a solution. For near free molecule or near continuum flows, a good initial approximation of $nf^{(0)}$ is immediately available. For the transition regime, the initial approximation may have to be found by first doing a cruder analysis, such as the Lees method discussed before. Several problems of the flow between two parallel plates are thus solved by this method (e.g., Willis [22]).

In applying to the shock structure problem, since the Navier-Stokes solution is reliable for the lower Mach numbers, and obtainable for any shock strength, it becomes an obvious choice as the initial approximation. Such was suggested by Burgers [3] in his analysis of the problem, but without actually carrying out the calculation. Recently Liepmann et al., [13], apparently independently, solved the problem by the same procedure in a computer, taking $\bar{\Theta} = p/\mu$, i.e., $\mathrm{Pr} = 1$. The solution shows no anomaly at least for Mach numbers as high as 10. A typical comparison against

the Navier-Stokes solution is schematically as shown. The agreement with the Navier-Stokes profile is very close in the downstream half, but the upstream portion is considerably more spread out, especially at the higher Mach numbers (see Figure 10). This is understandable since the effective coordinate is really $\int \bar{\theta}\, dx$, so the physical distance should be inversely proportional to the collision frequency, hence the density, which value for the upstream portion is a small fraction of that for the downstream portion for the stronger shocks.

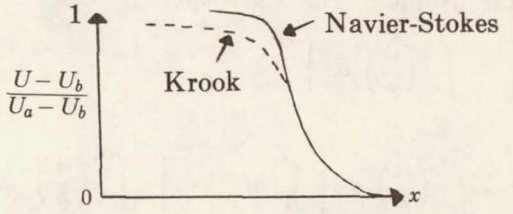


FIGURE 10. Comparison of Solutions

The Liepmann solution of the shock structure based on the BGK model is unquestionably the most satisfactory to date. It is, however, only the exact solution of an approximation to the Boltzmann equation.

IX. **Further discussion of the approximate solution of the Boltzmann equation.** Attractive as the BGK model is, we must not lose sight of the fact that it does not replace the Boltzmann equation. Better and better approximate methods presumably could be developed for the exact equation, and some of them would eventually surpass the BGK model in accuracy. Thus, we return to a discussion of the possible improvement in approximate methods, especially from the viewpoint that these should be applicable throughout the range from continuum to free molecule flow, as the BGK model is.

To gain some perspective, consider the simple differential equation

$$\epsilon \frac{df}{dt} = -f + f^{(0)}(t) \qquad f(0) \text{ given} \tag{IX.1}$$

where ϵ is a constant parameter of arbitrary magnitude.[2] Equa-

[2] I am indebted to Professor G. S. S. Ludford for calling attention to a very similar example in Erdelyi's paper on singular perturbations [5].

tion (IX.1) evidently embodies the most important features of the Boltzmann equation, ϵ having the same significance as $1/\overline{\Theta}$. It is also an analog of the BGK model. The boundary condition f(0) corresponds to the known distribution at an initial instant, or the boundary $s = s'$ as in Equation (VIII.6). The solution of Equation (IX.1) can be immediately written down,

$$f = f(0)\, e^{-t/\epsilon} + \frac{1}{\epsilon} e^{-t/\epsilon} \int_0^t f^{(0)} e^{t/\epsilon}\, dt \tag{IX.2}$$

which may be regarded as the simplified version of Equation (VIII.6).

If t is kept fixed and finite, the asymptotic solutions for $\epsilon \to \infty$ or $\epsilon \to 0$ are easily obtained from Equation (IX.2). For $\epsilon \to \infty$, the result is

$$f \cong f(0) + \frac{1}{\epsilon} \int_0^t [f^{(0)} - f(0)]\, dt + O\left(\frac{1}{\epsilon^2}\right), \tag{IX.3}$$

while for $\epsilon \to 0$, we get

$$f \cong f^{(0)} - \epsilon \frac{df^{(0)}}{dt} + O(\epsilon^2). \tag{IX.4}$$

The first term of Equation (IX.3) corresponds to the "free-molecule" flow approximation, and the second term is the equivalent of the "first collision" correction usually obtained by one iteration from the free-molecule solution. In the same analogy, the first term of Equation (IX.4) corresponds to the local Maxwellian in quasi-equilibrium, and the second term is the counterpart of the Chapman-Enskog correction. Both types of asymptotic solutions, as we now see, are *not uniformly valid for all t.* (Note, in particular, that Equation (IX.4) can never satisfy the prescribed boundary condition $f(0)$.) For a given ϵ however large, there is an upper limit of t beyond which the "free-molecule" type of asymptotic expansion ceases to be valid. In the other limit, for a given ϵ however small, there is a lower limit of t below which the asymptotic expansion for small ϵ is of no value. Although elementary, this demonstration seems to focus on some of the basic properties of the Boltzmann equation. The pitfalls of trying to push either type of asymptotic expansions into the "transition regime," where $\epsilon \sim O(1)$, are thus obvious.

The situations for $\epsilon \to \infty$ and $\epsilon \to 0$ are very similar to the problems of finding asymptotic solutions for low and high Reynolds numbers, resp., in viscous flow theory. Corresponding to the limit of $\epsilon \to \infty$, the "Stokes theory" for very low Reynolds numbers is known to be invalid in an unbounded fluid at sufficient distances from the body. In the other limit of $\epsilon \to 0$, the analogy to the Knudsen layer is the boundary layer near the body in conventional gasdynamics. The boundary layer thickness goes down as the viscosity is decreased. For points at fixed distances from the body surface, they will eventually lie in the effectively inviscid portion of the flow if the viscosity is small enough. To see the details in the boundary layer, the point in question must be made to move closer to the body surface as the viscosity is reduced, in order to remain within the boundary layer; then and only then the limit for vanishing viscosity may be taken. If we change the words "boundary layer" to "Knudsen layer" and "viscosity" to "mean free path," the last three sentences describe exactly what should be done for analysis of the Knudsen layer as $\epsilon \to 0$.

As also disclosed by the exact solution, Equation (IX.2), the natural independent variable should indeed be $\tilde{t} = t/\epsilon$. We now keep $\tilde{t}$ fixed and finite but let $\epsilon \to 0$. If Equation (IX.2) is expressed in $\tilde{t}$ and then integrated by parts, the resulting expansion is found to be

$$f \cong f(0)e^{-\tilde{t}} + \left[f^{(0)} - \epsilon \frac{df^{(0)}}{dt} \right] - e^{-\tilde{t}} \left[f^{(0)}(0) - \epsilon \frac{df^{(0)}}{dt} \bigg|_{t=0} \right] + O(\epsilon^2).$$

By expanding $f^{(0)}$ and $df^{(0)}/dt$ for small ϵ two alternative forms, both accurate to $O(\epsilon)$ *for finite* $\tilde{t}$, are obtained,

$$f \cong f(0)e^{-\tilde{t}} + \left[f^{(0)}(0) - \epsilon \frac{df^{(0)}}{dt} \bigg|_{t=0} \right] [1 - e^{-\tilde{t}}] + \epsilon \frac{df^{(0)}}{dt} \bigg|_{t=0} \tilde{t} + \cdots, \tag{IX.5}$$

$$f \cong f(0)e^{-\tilde{t}} + \left[f^{(0)} - \epsilon \frac{df^{(0)}}{dt} \right] [1 - e^{-\tilde{t}}] + \epsilon \frac{df^{(0)}}{dt} \bigg|_{t=0} \tilde{t}e^{-\tilde{t}} + \cdots. \tag{IX.5'}$$

The second form is clearly preferable since it remains valid as $\tilde{t} \to \infty$, merging smoothly into the "outer expansion" Equation (IX.4) and taking on the prescribed boundary value at $t = 0$.

It is now possible to estimate more accurately the "thickness" t_K of the Knudsen layer, in the sense that beyond which, to $O(\epsilon)$, Equations (IX.5′) and (IX.4) agree with each other. The condition is therefore

$$e^{-\tilde{t}_K} \sim O(\epsilon^2)$$

i.e.,

$$\tilde{t}_K \sim O(\ln \epsilon)$$

or

$$t_K \sim O(\epsilon \ln \epsilon).$$

In fact, no matter to what finite order of ϵ is expanded the asymptotic solution Equation (IX.4), the same argument will show that t_K is always $O(\epsilon \ln \epsilon)$. The Knudsen layer remains to be treated separately.

Returning to the central problem of formulating an approximate solution for the Boltzmann equation, we suggest that the distribution function should exhibit much the same basic features as the solution Equation (IX.2) of the simplified model. In a moment equation approach, for instance, a reasonable choice of the trial function might resemble Equation (IX.5′). If we assume an average collision frequency $\bar{\Theta}$ as in the BGK model, a convenient form is

$$nf = n'f' \exp\left[- \int_{s'}^{s} \left(\frac{\Theta}{\xi} \right) ds \right] + n_0 f_0 \left[1 - \exp\left[- \int_{s'}^{s} \left(\frac{\Theta}{\xi} \right) ds \right] \right] \tag{IX.6}$$

where $n_0 f_0$ is chosen to contain the adjustable parameters to be controlled by satisfying the moment equations. Since we work with only the average properties ρ, U_i, P'_{ij}, q_i in the moment equations, there is considerable leeway in the choice of $n_0 f_0$, the main restriction being that it must reproduce the Navier-Stokes and Fourier laws in the limit of $\bar{\Theta} \to \infty$. The free-molecule behavior is guaranteed by the $n'f'$ term, which automatically divides the molecules into groups depending on their "origin." No further

assumption such as the Lees "line of sight" principle is now necessary.

It may be noted that the term analogous to the last one in Equation (IX.5′) is omitted in Equation (IX.6) for brevity. The effect presumably is comparable to the net difference from alternative choices of $n_0 f_0$. It is to be emphasized, however, that Equation (IX.6) is not meant to be so assessed. Only the form is suggested by the BGK model. The approximation itself is adjusted to satisfy the moment functions of the exact Boltzmann equation, and any molecular model may be adopted for evaluating the collision integrals.

A source of difficulty in the use of Equation (IX.6) is the concept of an average collision frequency $\bar{\theta}$. As in the discussion of the BGK model following Equation (VIII.3), the choice of $\bar{\theta}$ appears to be either p/μ or $(2/3)(p/\mu)$. Besides this ambiguity, any choice of a single average $\bar{\theta}$ of course overestimates the mean free path of the fast-moving molecules, as shown by the smaller numerical factor 2/3 needed for matching the heat transfer by means of the BGK model. In assuming Equation (IX.6), on the other hand, we have effectively used the BGK model to suggest a way of grouping the free-molecule-like and Navier-Stokes-like molecules. Thus there is no strong reason not to allow $\bar{\theta}$ to vary somewhat with the speed of ξ, thereby compensating for this source of error. The refinement, however, may or may not be worthwhile, because, again, the net difference might be comparable to that from the alternative choices of $n_0 f_0$. In other words, the nature of the approximation Equation (IX.6) is, as a first step and like the BGK model, only to guarantee a smooth transition between the free-molecule and the Navier-Stokes limits. The resulting macroscopic equations are already rather cumbersome to attack, and have been solved only for the simple cases of the linearized plane and cylindrical Couette flows (Shen [18]). It seems yet premature to introduce further complications.

To conclude this brief survey of the current status of rarefied gasdynamics, we reiterate that our emphasis has been on the treatment of flow problems in terms of the observables such as mean velocity, pressure, temperature, shear stress and heat flux. The aim is thus essentially to look for the replacement of the conventional Navier-Stokes and Fourier relations in the hydrodynamic equations of motion, applicable throughout the entire range of Knudsen numbers. It might be said that to various degrees of approximation methods are indeed slowly emerging. Unfortunately

the geometry of the problem will always enter into resulting equations, so in effect special attention is required for each class of problems defined by its geometry. These equations furthermore are much more complicated than the Navier-Stokes, and our experiences are still confined to the simplest possible examples. The shock wave structure, because of its independence from solid boundaries, has been one of the ideal testing grounds for workers in this rapidly advancing field.

References

1. D. K. Ai, Calif. Inst. Tech. GALCIT Hypersonic Res. Proj. Memo. 56, (1960).
2. P. L. Bhatnagar, E. P. Gross, and M. Krook, Phy. Rev. **94** (1954), 511-525.
3. J. M. Burgers, Univ. of Md., Inst. Fluid Dyn. Appl. Math., Tech Note BN-83, 1956.
4. S. Chapman, and T. G. Cowling, *The mathematical theory of nonuniform gases*, 2nd ed. Cambridge Univ. Press, New York, 1952.
5. A. Erdelyi, Atti Accad. Sci. Torino **95** (1960-61), 651-672.
6. H. Grad, *On the kinetic theory of rarefied gases*, Comm. Pure Appl. Math. **2** (1949), 331-407.
7. _____, Comm. Pure Appl. Math **5** (1952), 257-300.
8. E. P. Gross, and S. Ziering, Phys. Fluids **1** (1958), 215-224.
9. J. Jeans, *The dynamic theory of gases*, 4th ed., Cambridge Univ. Press, New York and Dover, New York, 1925.
10. M. Krook, Phy. Rev. **99** (1955), 1896-1897.
11. L. Lees, Cal. Inst. Tech. GALCIT Hypersonic Res. Proj. Memo. 51, 1959.
12. L. Lees, and C. Y. Lin, *Rarefied gas dynamics*, 2nd Sympos., Academic Press, New York, 1961, pp. 391-428.
13. H. W. Liepmann, R. Narasimha, and M. T. Chahine, Phys. Fluids **5** (1962), 1313-1324.
14. H. M. Mott-Smith, *The solution of the Boltzmann equation for a shock wave*, Phys. Rev. **82** (1951), 885-892.
15. C. Muckenfuss, Phys. Fluids **3** (1960), 320-321.
16. P. Rosen, *Use of restricted variational principles for the solution of differential equations*, J. Appl. Phys. **25** (1954), 336-338.
17. A. Sakurai, *A note on Mott-Smith's solution of the Boltzmann equation for a shock wave*, J. Fluid Mech. **3** (1957), 255-260.
18. S. F. Shen, *Rarefied gas dynamics*, Third Sympos. Vol. II, Academic Press, New York, 1963, pp. 112-131.
19. F. S. Sherman, and L. Talbot, *Rarefied Gas Dynamics*, 1st Sympos. Pergamon Press, London, 1960, pp. 161-191.
20. H. S. Tsien, J. Aero. Sci. **13** (1946), 653-664.
21. D. R. Willis, Princeton Univ. Aero, Eng. Dept. Rept. 442, 1958.
22. _____, Phys. Fluids **5** (1962), 127-135.
23. S. Ziering, F. Ek, and P. Koch, Phys. Fluids **4** (1961), 975-987.
24. K. Zoller, Z. Phys. **130** (1951), 1-38.

CORNELL UNIVERSITY

F. K. Moore

Models of Gas Flows with Chemical and Radiative Effects

I. **Differential equations and basic models.** In the course of this chapter we will consider flows with chemical activity and in which radiative effects are of significance and, in particular, those pertaining to re entry and propulsion. We will be concerned mainly with chemical effects, and will not go into detail in connection with electrical (ionization) effects.

a. *Entry Phenomena.* Briefly, the hypersonic entry of a vehicle into an atmosphere may be described as follows: Initially, as the vehicle begins to penetrate the very rarefied outer atmosphere, it is subjected to a bombardment by the gas particles in its path. In the region where the mean free path of the molecules is very large compared with the dimensions of the body, the rebounding molecules do not interfere with other approaching molecules, and the vehicle suffers only the retarding effect due to direct collisions with the particles in its path. Further penetration into the denser regions of the atmosphere results in the establishment of a "flow field," characterized by a mean free path somewhat less than the characteristic dimension of the vehicle. Thus, rebounding molecules encounter other molecules in the region surrounding the vehicle, so that the particles in the path of the vehicle are, to some extent, warned of its approach. As still lower altitudes are reached, these warning signals

coalesce into a strong shock wave standing ahead of the vehicle.

Atmospheric entry, then, involves a progression from free molecular flow to continuum flow. In the region of the shock, collisions promote excitation of higher levels of internal energy of the molecules (vibration and rotation) resulting in first, dissociation, and second, if the velocity is sufficiently high, ionization of the gas. The flow field is shown schematically in Figure 1. Such flow fields of highly excited air are studied to determine rates of heat transfer at the surface of the vehicle. In their study, one must consider the behavior of the "real" gas at elevated temperatures. Furthermore, the nature of the flow field governs the mechanics of the motion of the vehicle, the drag being of particular interest. One also finds that ionization of the flow field affects communications with the vehicle. Thus, the study of high-temperature gas is of great importance.

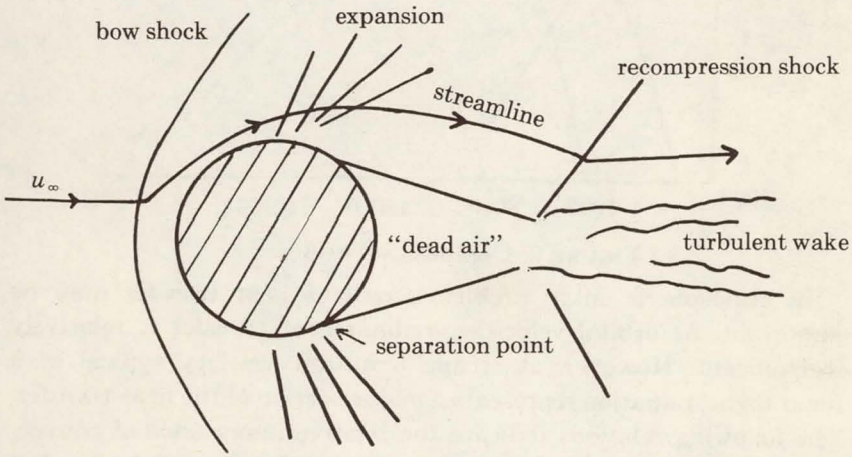


FIGURE 1. Aerodynamic Phenomena at High Speed

As the temperature of a gas is increased on passing through a shock, compositional changes will take place. Considering air, we have the following dissociation processes:

$$O_2 \rightleftharpoons O + O$$

$$N_2 \rightleftharpoons N + N$$

as well as those involving NO formation, ionization, and others.

Figure 2 shows the concentrations of air versus temperature, at equilibrium. We note from the figure that oxygen shows a marked increase in dissociation at about 4000°K, whereas nitrogen does not dissociate appreciably until temperatures in excess of 8000°K are reached. 8000°K corresponds to velocities greater than those encountered in earth orbits, such high speeds being characteristic of entry from a lunar trajectory.

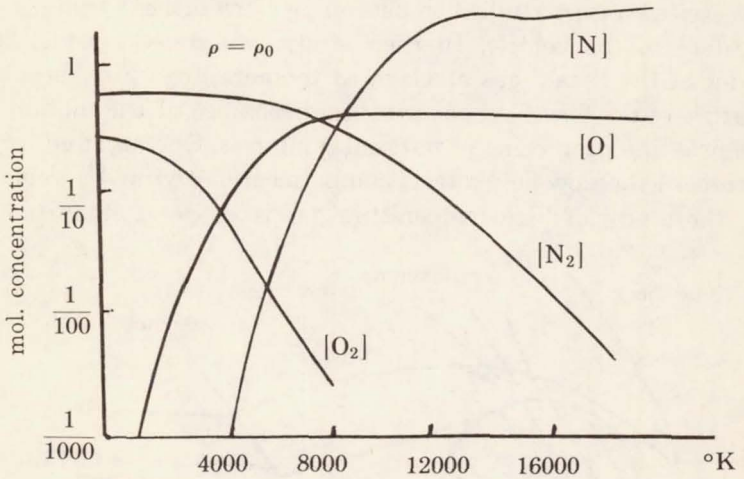


FIGURE 2. Composition of Air

In atmospheric entry problems, radiant heat transfer may be important. At orbital velocities, radiant heat transfer is relatively insignificant. However, at escape (capture) velocity typical of a lunar flight, radiation represents a major portion of the heat transfer. The following relations indicate the relative importance of convective and radiative transfer, and the dependence upon velocity: Let q be the heat load to the vehicle (see [1]). Then:

$$q_{\text{conv}} \propto (\rho/L)^{1/2} U_\infty^3,$$

$$q_{\text{rad}} \propto \rho^{3/2} L U_\infty^{10},$$

where ρ is the density, L is a characteristic body dimension, and U_∞ is the vehicle velocity. The very strong dependence of radiative transfer on velocity is to be expected, because $q_{\text{rad}} \propto T^4$ and $T \propto U_\infty^2$. Thus, one finds that at 25,000 ft/sec. q_{rad} is approximately 10% of the heat load, while at 35,000 ft/sec. q_{rad} is the dominant heat load factor.

b. *Nonequilibrium.* The composition of the gas at any point in the flow field is, of course, dependent upon the chemical kinetics of the gas. If the gas chemistry does not have time to equilibrate, that is to say, if the composition is not the same as the equilibrium composition for the local temperature, one has a nonequilibrium flow. At very high speeds and very high altitudes, flows may be dominated by nonequilibrium effects.

Of course, in a general sense, all flows are "nonequilibrium" situations. In basic fluid dynamics, the Reynolds number

$$\mathrm{Re} = \rho U_\infty L/\mu$$

is a governing parameter, where U_∞ is a characteristic velocity, L is a characteristic length, ρ is the density and μ the viscosity. Now, the Reynolds number is in fact a comparison of relaxation time with the time of passage of flow. Consider the time required for a diffusive process to occur, i.e., the relaxation time:

$$\tau_{\text{relax}} = \rho L^2/\mu.$$

For very large distances, a very long relaxation time is required, similarly, diffusive effects are slow for very small viscosity. Now, compare τ_{relax} with the time of passage of the flow over a body of characteristic dimension L:

$$\tau_L = L/U_\infty.$$

The ratio of these two characteristic times is

$$\frac{\tau_{\text{relax}}}{\tau_L} = \frac{\rho L U_\infty}{\mu} \equiv \mathrm{Re}. \tag{1}$$

This ratio is just the Reynolds number of the flow, and compares the time for a mixing process to occur with the time of flow passage over the body.

Similar parameters appear in other flows, and by way of comparison, we note that using the time of diffusion of a magnetic field yields

$$\frac{L^2 \mu_m \sigma}{L/U_\infty} = L U_\infty \mu_m \sigma \equiv \mathrm{Rm},$$

where Rm is the so-called magnetic Reynolds number. The ratio of chemical relaxation time to the time of passage yields the following:

$$\frac{\tau_{\text{chem}}}{\tau_L} = \frac{\tau_{\text{chem}}}{L/U_\infty}. \tag{2}$$

This may be of unit order in high-speed flows at high altitudes.

For ordinary, viscous hydrodynamic flows, $\text{Re} \gg 1$, i.e., the diffusive relaxation time is much greater than the time of passage. In this case the flow may be said to be dynamically "frozen," there is insufficient time for the decay process to reach equilibrium. If the chemical parameter $\tau_{\text{chem}}/\tau_L \gg 1$, we may speak of chemical freezing. Now, if the ratio $\tau_{\text{chem}}/\tau_L = O(1)$ one encounters serious difficulties in analysis, just as in ordinary hydrodynamic flows one encounters analytical problems when the Reynolds number is of order unity. On the other hand, if the ratio is much less than one (equilibrium flow), great simplifications result.

The study of chemically reacting flows is, as one would expect, considerably more complicated than nonreacting flows. In the latter, one finds similarities from dimensional analysis, and these can be used to great advantage. In chemical kinetics, however, such similarities do not generally occur, and therefore one must attempt to solve particular problems and hope to find simplifications which will render the analysis tractable (see [2]).

c. *Equations of Motion.* From basic fluid dynamics we have the continuity equation:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0. \tag{3}$$

Since the gas consists of several constituents we require a continuity equation for the separate constituents:

$$\rho \frac{Dc_i}{Dt} - W_i = -\frac{\partial}{\partial n}(\rho c_i u_{n_i}), \tag{4}$$

where

$$\frac{D}{Dt} \equiv \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla$$

and the other symbols are:

c_i, the mass fraction of the ith constituent;
ρ, the density;
W_i, a production term to be discussed shortly;

$\partial/\partial n$, the gradient normal to the surface;

u_{n_i}, the diffusion velocity of the ith constituent.

Equation (4) states that the rate of increase of the particular species is equal to the rate of production (by chemical reaction) of that species, minus the diffusion of that species across some control surface. We note that:

$$\sum_i c_i = 1.$$

The bracketed quantity on the right-hand side of Equation (4) deserves attention. We can write:

$$\rho c_i u_{n_i} = \rho m_i \left(\sum_k \frac{c_k}{m_k}\right)^2 \sum_j m_i D_{ij} \left\{ \frac{\partial}{\partial n} \left(\frac{c_j}{\sum_k \frac{m_j}{m_k} c_k} \right) + \left[\frac{\sum_k c_k \left(1 - \frac{m_j}{m_k}\right)}{\sum_k \frac{m_j}{m_k} c_k} \right] \frac{\partial \ln p}{\partial n} \right\}$$

$$+ D_i^T \frac{\partial \ln T}{\partial n}.$$

The terms here are:

D_{ij}, diffusion coefficient for inter-diffusion of constituents;

$\partial \ln p/\partial n$, the "pressure diffusion," i.e., due to grad p; (In most flowfields $\partial p/\partial n$ may be neglected here.)

$D_i^T(\partial \ln T/\partial n)$, thermal diffusion, diffusion due to temperature gradient only; this is usually neglected, since the coefficient is small.

Thus, u_{n_i} is seen to depend largely upon the concentration gradient. The equation of state is:

$$p = \rho \left(\sum_i \frac{c_i}{m_i}\right) RT, \tag{5}$$

where R is the universal gas constant. The momentum equation is:

$$\rho \frac{D\mathbf{v}}{Dt} + \nabla p = \frac{\partial}{\partial n} \left(\mu \frac{\partial \mathbf{v}}{\partial n} \right). \tag{6}$$

The energy equation is:

$$\rho \frac{DH}{Dt} - \frac{\partial p}{\partial t} - \mathbf{v} \cdot \nabla p - q_R = \mu \left(\frac{\partial u}{\partial n}\right)^2 + \frac{\partial}{\partial n} \left[\lambda \frac{\partial T}{\partial n} - \rho \sum_i u_{n_i} H_i \right]. \tag{7}$$

In these, we identify the following quantities:

H, specific enthalpy, and $H = \sum_i H_i = \sum_i c_i(h_i + h_i^{(0)})$, where h_i is specific internal enthalpy and $h_i^{(0)}$ is the (potential) chemical energy of dissociation;

q_R, a heat source term, for example, radiation heating;

$\mu(\partial u/\partial n)^2$, the viscous dissipation term;

$\lambda \partial T/\partial n$, the heat conduction term;

$\sum_i u_{n_i} H_i$, the diffusion of enthalpy.

d. *Lumped Constituents.* A useful approximation, or gas model, frequently employed is that of "lumped constituents." In effect, we normally use the lumped constituents idea in aerodynamic problems, letting air take on an average molecular weight, etc., neglecting interdiffusion of constituents as well as chemical reaction. For low-temperature problems, it is not necessary that the constituents be similar.

In cases where chemistry is involved, we can use the same approach to simplify the analysis. Briefly, for air, we have a mixture of O_2 and N_2, and in dissociating flows we have, in addition, O and N. Let us combine these four constituents, considering the mixture of O and N as a monatomic gas which we denote by A, and the mixture of O_2 and N_2 as a diatomic gas, which we denote by A_2. We then have a two-constituent gas, and the i's of Equations (4) to (7) may have the values 1, 2. Now, this "lumping" requires that the molecular weights of the two gases be nearly the same, as well as that there be no net diffusion or reaction between gases lumped in the same group. That is, we assume that the interdiffusion of O and N is small, and that the reactions between O and N are negligible, and similarly for the "A_2" portion of the gas. It is also necessary that the heats of dissociation be nearly the same for both reacting pairs. Thus, we shall consider reactions of the form:

$$A_2 \leftrightharpoons A + A.$$

and the right-hand side of Equation (4) becomes:

$$\rho c_i u_{n_i} = \rho D \frac{\partial c_i}{\partial n}.$$

e. *Binary Mixture.* We will now use the subscript 1 to denote A atoms, and 2 to denote A_2 molecules. The equation for specific enthalpy may then be written:

$$H = \frac{RT}{m_2}\left\{5c_1 + 4c_2\left[\frac{7}{8} + \frac{1}{4}\left(\frac{T_v/T}{e^{T_v/T} - 1}\right)\right]\right\} + c_1 h_1^{(0)}, \tag{8a}$$

where $5Rc_1/m_2$ is the specific heat at constant pressure of a monatomic gas, $7Rc_2/2m_2$ is that of a rotationally-excited diatomic gas, the term $c_2 T_v/T(e^{T_v/T} - 1)$ accounts for the energy of vibration, and finally, $c_1 h_1^{(0)}$ is the energy of dissociation (see [5]). T_v is the characteristic temperature for vibration. Since T_v is about 3000° K for air, the quantity $T_v/4T(e^{T_v/T} - 1) \approx 0$ for air at room temperature. Also, we note that

$$0 < \left(\frac{T_v/T}{e^{T_v/T} - 1}\right) < 1.$$

f. *Lighthill Ideal Dissociating Gas.* The foregoing gas model is not calorically perfect, for Equation (8a) indicates that the specific heat depends upon temperature. In [5], Lighthill observed that the bracketed quantity above was nearly constant for a large range of temperature and he assumed that $T_v/T(e^{T_v/T} - 1) = \frac{1}{2}$ (midway between the extremes 0 and 1) thus rendering the model calorically perfect. This is the first Lighthill gas approximation, and with it we have:

$$H = \frac{RT}{m_2}(4 + c_1) + c_1 h_1^{(0)} \tag{8b}$$

which is the relationship for the enthalpy of an ideal, calorically perfect, dissociating gas. A further approximation due to Lighthill is as follows: Consider the equilibrium concentrations of the atoms and molecules, according to the law of mass action,

$$\left(\frac{c^2}{1 - c}\right)_{eq} = \frac{\rho_D}{\rho} e^{-(2h^{(0)}/(RT/m_2)}, \tag{9}$$

where c is the atom concentration. From this equation it is seen that at high altitudes, i.e., low densities, one could expect to find higher concentrations of atoms. The characteristic density is actually dependent upon temperature:

$$\rho_D \propto \sqrt{(T)}(1 - e^{T_v/T}).$$

For temperatures of interest, ρ_D exhibits a relative maximum, and is fairly constant, and is therefore assumed constant, without introducing serious error. In fact, we may add to Lighthill's argument the

observation that ρ_D reaches a maximum at precisely the temperature

$$T = \frac{2T_v}{e^{T_v/T} - 1}$$

This is the temperature at which Equation (8b) is exact; i.e., the square bracket of Equation (8a) for specific enthalpy is just equal to unity. Thus, Equations (8b) and (9) are consistent descriptions of a dissociating gas which is just 50% excited vibrationally.

The foregoing "Lighthill model" is not particularly powerful analytically, though T_v is eliminated as a parameter. It does not result in any appreciable simplification in machine calculations. The main value of the model has been that it provides a standard dissociating gas in terms of which various computations and theories may be compared.

g. *Dissociation Kinetics.* Consider the reaction

$$\mathrm{M} + \mathrm{A}_2 \underset{k_D}{\overset{k_R}{\rightleftharpoons}} \mathrm{A} + \mathrm{A} + \mathrm{M},$$

where M may be A_2 or A. M is the third party to the process and is the particle which, in collision, shares the energy of either dissociation or recombination. We can write

$$(10)\quad \frac{W_1}{\rho} = \text{const}\frac{T_p^{-(s+1)}}{1+c}\left[1 - c + 2\,\frac{k_D^{(1)}}{k_D^{(2)}}c\right]\left[(1-c)e^{-T_D/T} - \frac{\rho}{\rho_D}c^2\right],$$

where W_1 is the production rate for atoms (zero for equilibrium), and $T - s$ is a temperature dependence factor. This equation embodies the relation between k_D and k_R at equilibrium ($W_1 = 0$) obtained from the mass action law. The first square bracket accounts for the different third body in the collision, i.e., whether M is A or A_2. The term $(1 - c)e^{-T_D/T}$ in the second bracket deals with the forward reaction (dissociation), while the term $-(\rho/\rho_D)c^2$ deals with the reverse reaction (recombination). Substituting for ρ/ρ_D from Equation (9) renders the second square bracket zero, which is the result desired for equilibrium. It is important to note that the dissociation in term Equation (10) is second order in density, while the recombination term is third order in density.

In addition, it is noted in [3] and [4] that from experiment:

$$\frac{k_D^{(1)}}{k_D^{(2)}} \approx 35\,\frac{T}{T_D},$$

Freeman, in a paper emphasizing the Lighthill model, has, in effect, taken the first square bracket to be equal to a constant. This simplification of Equation (10) is therefore often taken as one of the specifications of Lighthill's ideal gas.

II. **Sound waves.**

a. *Perturbation Equations for Chemical Nonequilibrium.* For sound waves (acoustics) with chemical relaxation and radiative effects, we first write the perturbation equations for a binary mixture of gases. The primed quantities denote the perturbations, e.g.,

$$\rho = \rho_0 + \rho', \quad p = p_0 + p', \quad \mathbf{v} = \mathbf{v}'.$$

The linearized equations are then:

Continuity: $$\frac{\partial \rho'}{\partial t} + \rho_0 \nabla \cdot \mathbf{v}' = 0$$

Momentum: $$\rho \cdot \frac{\partial \mathbf{v}'}{\partial t} + \nabla p' = 0$$

(11) Energy: $$\rho \cdot \frac{\partial H'}{\partial t} - \frac{\partial p'}{\partial t} = 0$$

State: $$\frac{p'}{p_0} = \frac{c'}{1 + c_0} + \frac{\rho'}{\rho_0} + \frac{T'}{T_0}$$

Production Equation: $$\frac{W_1}{\rho_0} \propto \left[(1 - c) e^{-T_D/T} - \frac{\rho}{\rho_D} c^2 \right].$$

From Equation (9) $[c^2/(1 - c)]_{eq} = (\rho_D/\rho) e^{-T_D/T}$, we substitute in Equation (10) for ρ/ρ_D and obtain

$$\frac{W_1}{\rho} \propto \left[(1 - c) e^{-T_D/T} - \frac{c^2(1 - c_{eq})}{c_{eq}^2} e^{-T_D/T} \right]$$

which, after factoring, including $[((1 - c_{eq})(1 - c)/c_{eq}^2) e^{-T_D/T}]$ in a proportionality factor, and changing sign, gives

(12) $$-\frac{W_1}{\rho} \propto \left[\frac{c^2}{1 - c} - \frac{c_{eq}^2}{1 - c_{eq}} \right].$$

All the quantities lumped into the proportionality factor are functions of state. Now, if equilibrium prevails, $c = c_{eq}$ and we have $W_1 = 0$. If $c \neq c_{eq}$, the expression indicates that the production rate will vary, to cause the mixture to tend toward equilibrium. In this

derivation we express the mass fraction as $c = c_0 + c'$ and assume that c' is a small quantity. We also note that $c_{eq} = c_0 + c'_{eq}$, where c_0 is the free-stream equilibrium concentration, to account for the dependence of c_{eq} upon local temperature and density. From the proportionality, Equation (12), we can linearize the term

$$\frac{c^2}{1-c} - \frac{c_{eq}^2}{1-c_{eq}}$$

to obtain $c_0(2-c_0)(c'-c'_{eq})/(1-c_0)$ and finally write

$$-\frac{W_1}{\rho_0} = \frac{1}{\tau}(c' - c'_{eq}) = -\frac{\partial c'}{\partial t}. \tag{13}$$

Here all the proportionality terms are lumped into τ, which is a relaxation time.

We now generate a sound wave within the gas and examine its effects. If c'_{eq} were constant, one would simply solve Equation (13) to find:

$$c' = c'_{eq}(1 - e^{-t/\tau}).$$

This is a relaxation equation, where the concentration goes from an initial to a final level following a simple exponential curve. However, since c'_{eq} is a function of temperature and density, such a solution is rarely of value.

We now specify, as in ordinary acoustics, that the wave be curl-free, so that:

$$\mathbf{v}' \equiv \nabla\phi,$$

where ϕ is the potential. The acoustic equation obtained from Equations (11) and (13) is now

$$\tau^* \frac{\partial}{\partial t}\left(\frac{\partial^2\phi}{\partial t^2} - a_f^2\nabla^2\phi\right) + \frac{\partial^2\phi}{\partial t^2} - a_e^2\nabla^2\phi = 0. \tag{14}$$

(See [6], [7], [11], [12], [13], and [14].)

The second parenthesis is the left side of the usual acoustic wave equation, with a_e, the equilibrium sound speed (isentropic) given by:

$$a_e \equiv \sqrt{(\gamma RT)}$$

(R is the universal gas constant divided by the molecular weight), τ^* is a reference relaxation time, and if τ^* is large, we have frozen flow, since the first bracket predominates. The first bracket is the frozen wave operator, wherein the velocity of propagation is a_f, the

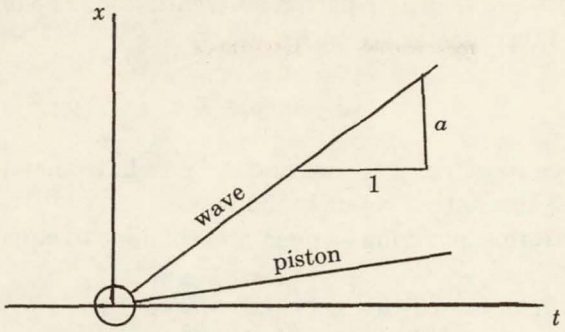


FIGURE 3. Piston and Wave Motions

frozen sound speed. a_f is found to be somewhat greater than a_e, and one may think of the gas as being stiffer in the frozen flow case. If the relaxation time is short, i.e., τ^* small, we may neglect the first bracket, and we have ordinary equilibrium flow, with acoustic propagation of small disturbances at the speed a_e.

If τ^* is of the order of the period of the disturbance, the equation can be solved exactly, but the solution is not particularly edifying. It is of more interest to begin by considering a wave with a discontinuity, i.e., a jump wave as produced by a piston. A general solution of the classical wave equation is $\phi = \phi(x - at)$, which includes the step function $\phi = 1(x - at)$. We represent the piston and wave motions on a t-x plot, as in Figure 3.

In the case of combined waves subject to Equation (14), one expects a solution closely related to that for the classical wave equation. Accordingly, we transform to coordinates ξ and η, where

$$\xi \equiv \frac{1}{a_f \tau^*}(x - at)$$

(15) and

$$\eta \equiv \frac{1}{a_f \tau^*} x.$$

Thus, ξ is the distance measured from a characteristic as yet undefined, and η is the distance the piston has traveled. Equation (14) becomes

$$-\frac{a}{a_f}\left[\frac{a^2}{a_f^2}\phi_{\xi\xi\xi} - (\phi_{\xi\xi\xi} + 2\phi_{\xi\xi\eta} + \phi_{\xi\eta\eta})\right]$$

(16)

$$+\left[\frac{a^2}{a_f^2}\phi_{\xi\xi} - \frac{a_e^2}{a_f^2}(\phi_{\xi\xi} + 2\phi_{\xi\eta} + \phi_{\eta\eta})\right] = 0,$$

where the subscripts denote partial differentiation. The highest order derivatives must vanish

$$\frac{a^2}{a_f^2}\phi_{\xi\xi\xi} - \phi_{\xi\xi\xi} = 0$$

and hence we must have $a = a_f$, and we conclude that jump-waves must propagate at the frozen sound speed.

Now, the terms involving $\phi_{\xi\xi}$ must also combine to equal zero, thus:

$$-\frac{a}{a_f}(-2\phi_{\xi\xi\eta}) + \frac{a^2}{a_f^2}\phi_{\xi\xi} - \frac{a_e^2}{a_f^2}\phi_{\xi\xi} = 0$$

and because $a = a_f$,

$$2\phi_{\xi\xi\eta} + \left(1 - \frac{a_e^2}{a_f^2}\right)\phi_{\xi\xi} = 0.$$

Upon integrating once, along the wave front, we find

$$(\phi_{\xi\xi})_{\xi=0} = e^{-\epsilon\eta} \tag{17}$$

where

$$\frac{1}{2}\left(1 - \frac{a_e^2}{a_f^2}\right) \equiv \epsilon$$

is a rather small quantity. This equation shows that the jump amplitude decreases with distance, (i.e., $\eta \propto x$) and we have a decay of amplitude of the wave head, as illustrated in Figure 4, which is taken from [7].

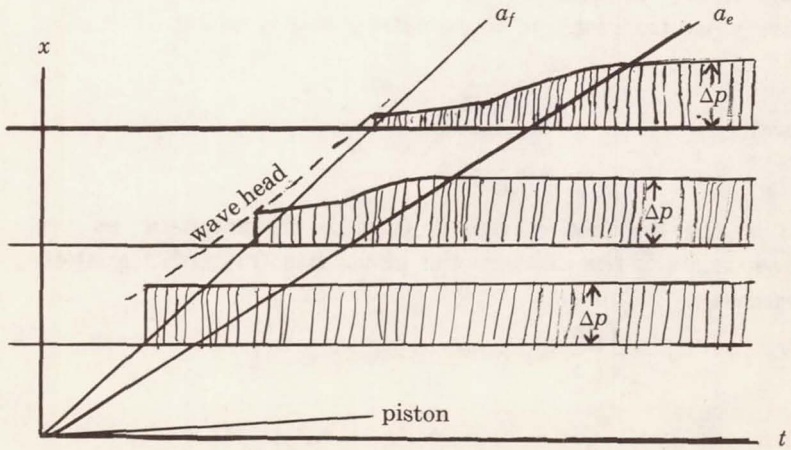


FIGURE 4. Decay of Sound Wave

b. *The Telegraph Equation.* Allowing ϵ to approach zero, and redefining the piston-travel coordinate by $\epsilon\eta \equiv \zeta$, we get a new equation, after integrating Equation (16) once with respect to ζ:

$$\phi_{\xi\zeta} + \phi_\xi - \phi_\zeta = 0 \tag{18}$$

for small ϵ. This Equation (18) is related to the telegraph equation, and provides a model for the relaxing sound wave. Equation (18) can be solved exactly:

$$v' = e^{-(\xi+\zeta)} \frac{\partial}{\partial \xi} \int_0^\xi I_0(2\sqrt{\{\zeta[\xi - \omega]\}})\, e^{\omega} f(\omega)\, d\omega,$$

where I_0 is a zero-order Bessel function of an imaginary argument, and $f(\omega)$ is a source function for the piston motion. The nature of the solution indicates how the frozen wave decays and changes to an ordinary wave. The decay is due to energy absorption in dispersing the wave. Thus, when chemical activity is present, and we have non-equilibrium conditions, there results an interplay between the chemical kinetics and the dynamic processes. This results in the dispersion of the wave. The relaxation time plays a role somewhat analogous to that of viscosity.

c. *Waves Affected by Heat Sources.* Turning now to the acoustic problem involving heat sources which might be thought of as due to radiative heat transfer, we write the equations of motion for one dimension:

$$\begin{aligned}
&\text{Continuity:} && \frac{\partial \rho'}{\partial t} + \rho_0 \frac{\partial u'}{\partial x} = 0 \\
&\text{Momentum:} && \rho_0 \frac{\partial u'}{\partial t} + \frac{\partial p'}{\partial x} = 0 \\
&\text{Energy:} && \rho_0 \frac{\partial H'}{\partial t} - \frac{\partial p'}{\partial t} \equiv q'(T') \\
&\text{State:} && \frac{p'}{p_0} = \frac{\rho'}{\rho_0} + \frac{T'}{T_0}
\end{aligned} \tag{19}$$

and we note that the right-hand side of the energy equation is a heat source term and plays somewhat the same role as does the chemical production term in Equations (11). We also have $u' = \phi_x$, $p' = -\rho\phi_t$ as before. Manipulation of these equations yields

$$\phi_{tt} - a^2\phi_{xx} = -\frac{\gamma - 1}{\rho_0}\, q'(T'), \tag{20}$$

where a is the ordinary isentropic sound speed. A second equation also results, which is

$$\phi_{tt} - \frac{a^2}{\gamma}\phi_{xx} = -R\frac{\partial T'}{\partial x}, \tag{21}$$

where a^2/γ is the "isothermal" sound speed: For an isothermal wave, $a_T^2 = (\partial p/\partial\rho)_T = RT$, and thus $a^2/\gamma = a_T^2$.

The problem now is to combine Equations (20) and (21). If $q' = q'(T')$, then, in principle, T' may be eliminated between Equations (20) and (21). In order to illustrate the effects of radiative transfer, whereby the hot region loses heat to the cold region, we will be more specific, and write the simple proportionality

$$q' = -\left[\frac{1}{k}\frac{\rho_0 R}{\gamma - 1}\right]T'.$$

Then, elimination of T' yields Equation (14) again, except that τ^* is replaced by k. Thus,

$$k\frac{\partial}{\partial t}\Box_s^2\phi + \Box_T^2\phi = 0, \tag{22}$$

where $\Box^2$ is the wave operator, $\partial^2/\partial t - a^2(\partial^2/\partial x^2)$, the subscripts s and T being for the isentropic and the isothermal sound speeds respectively. If k is very large, we get the isentropic wave. If k is very small we get the isothermal wave. For cases where radiation is intense, the effect of the radiative heat transfer will generally be to redistribute the energy and reduce temperature gradients. We note at this point that a^2 is only slightly greater than a^2/γ since γ is near unity, and one could also derive a form of the telegraph equation (Equation (18)) for this heat addition case.

d. *Relation to Radiation Transport.* The foregoing assumption, that $\partial q'/\partial T' = \text{const}$ is not actually valid for radiative transport, and we must examine more fully the transport of energy by radiation to determine the proper general expression for q_R. (See [8] and [9].) Radiative energy emitted by an element of the hot gas may be absorbed by another element, at some distance, or "penetration depth" from the first. This distance is expressed as a reciprocal absorption coefficient, $1/\alpha_\nu$ for a given frequency ν. For a gas, $1/\alpha_\nu$ could be as long as

one kilometer. If the gas is in a uniform state over distances much larger than this, full "radiative equilibrium" prevails, and there is no heat flux, because the heat emitted and absorbed by each element is the same. However, in problems of interest in hypersonic flows, one must consider the hot gases at the nose to be confined to a small region, only centimeters thick. We cannot therefore, assume full radiative equilibrium (see Figure 5). In flow about such bodies, the path length of radiation is large compared to a characteristic dimension for temperature change in the flow. It is necessary then, to evaluate the integral,

$$q_R = \int_0^\infty (A_\nu - E_\nu) d\nu, \tag{23}$$

where A_ν is radiant energy absorbed, and E_ν is energy emitted, in the frequency range ν to $\nu + d\nu$. The theory to be outlined is in *Radiative transfer* by Chandrasekhar [8], and is reviewed also by Lighthill [9], Goulard [10], and Vincenti and Baldwin [11].

e. *Quasi-equilibrium Assumption.* In dealing with Equation (23), it is commonly assumed that atoms and molecules are in local thermal equilibrium, so that the gas element emits as a black body. This requires that particle collisions be much more frequent than photon emissions. Then one may write

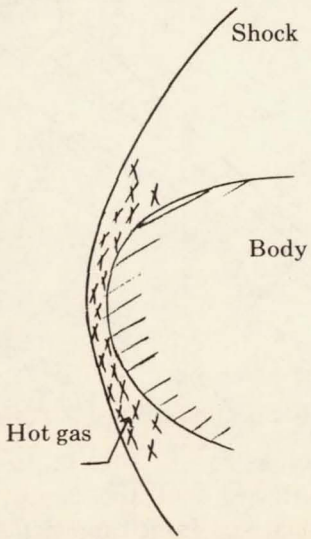


FIGURE 5. Nose-Cap Region

$$E_\nu = 4\pi\alpha_\nu B_\nu,$$

where B_ν is the black-body energy flux obtained from the statistical mechanics of a "photon gas," and $\int_0^\infty B_\nu d\nu = (1/\pi)\sigma T^4$, σ being the Stefan-Boltzmann constant. Now, by the "quasi-equilibrium" assumption the absorption coefficient α_ν is also taken to have its black-body value, not only for emission, but for absorption as well:

$$A_\nu = \alpha_\nu \int_0^{4\pi} I_\nu(\Omega)\, d\Omega. \tag{23a}$$

For full radiative equilibrium, $I_\nu = B_\nu/4\pi$, but here we must imagine that the "spectral intensity," I_ν, results from emission somewhere else at some other temperature. Ω is the solid angle defining the direction of the incoming radiation (see Figure 6). I_ν may be found from the "equation of radiative transfer" (Chandrasekhar [8, p. 9]).

$$-\frac{\partial I_\nu}{\partial \delta} = \alpha_\nu(I_\nu - B_\nu) \tag{24}$$

which says that along its path, s, the intensity diminishes by absorption $(\alpha_\nu I_\nu)$ and is augmented by black-body emission $(\alpha_\nu B_\nu)$, scattering into or out of the beam being neglected.

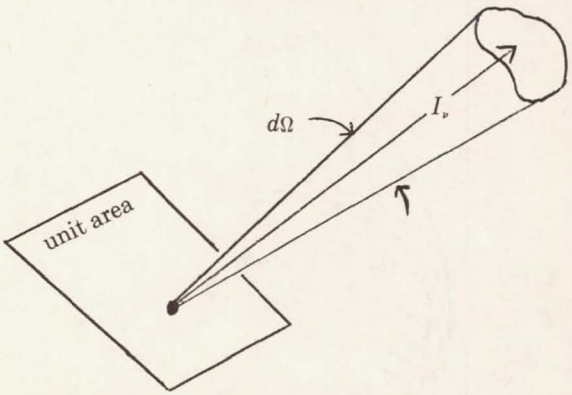


FIGURE 6. Spectral Intensity

Equation (24) may be solved for I_ν, subject to boundary conditions about bounding surfaces, and the result, via Equation (23a), may be used to evaluate q_R from Equation (23). In carrying out this process, a constant value of α_ν is commonly assumed.

This disregard of the frequency dependence of α is called the "gray-gas" assumption. (This has been recently described in [11].)

f. *The Rosseland Limit.* We have mentioned that $1/\alpha$ may be as large as 10^5 cm. A characteristic length, say L, is likely to be much larger than this (radiative equilibrium) only in astrophysical problems. If $1/\alpha$ is small (but not negligibly so) compared with L, then it may be regarded as a photon free path length, and is this "Rosseland limit," radiative transfer depends upon grad T, just as heat transfer by conduction does. In fact, Rosseland found,

$$q_R = \frac{16}{3}\frac{\sigma T^3}{\alpha}\text{grad } T.$$

The foregoing Rosseland formula cannot be used for shock layers because $L \ll 1/\alpha$, in general. However, the quasi-equilibrium assumption is usually quite good. Ordinarily, there are, say 10^{10} collisions per second for particles, and $c\alpha$ is $3 \times 10^{10}/10^5$ or only about 10^5 photon interactions per second, and thus, the requirement is met. At very high altitude, the collision frequency would be too low to maintain quasi-equilibrium, and radiation transfer would be "collision limited."

Radiation pressure and the contribution of radiation to the internal energy of the gas are usually neglected. The following example will serve to indicate the magnitude of these effects and show why it is reasonable to omit them. Consider the intensity of black-body radiation from a source at 8000°K. The energy flux E_R is given by

$$E_R = \frac{1}{\pi}\sigma T^4 \approx \frac{5.67 \cdot 10^{-8} \cdot 4.1 \cdot 10^{15}}{3 \cdot 14} \approx 8 \cdot 10^3 \frac{\text{joule}}{\text{cm}^2\text{sec}}.$$

To find the corresponding specific density E_D, i.e., the internal energy of radiation, we divide by the velocity of light, c,

$$E_D \approx \frac{8 \cdot 10^3}{3 \cdot 10^{10}} \approx 3 \cdot 10^{-7}\frac{\text{joule}}{\text{cm}^3} = 3\frac{\text{erg}}{\text{cm}^3}$$

and the radiation pressure is

$$p_R = \frac{1}{3}E_D \approx 1\frac{\text{dyne}}{\text{cm}^2}.$$

Thus, while the energy transfer by radiation is considerable, the internal energy due to radiation, and radiation pressure, are both negligible when compared to the enthalpy of the order of $10^8\,\mathrm{erg/cm^3}$, and the static pressure of the order $10^5\,\mathrm{dyne/cm^2}$, which are typical of hypersonic flows.

g. *Application of Radiative Transfer Theory to Waves.* The problem of the effect of thermal radiation on acoustic waves has been investigated in [12], and quite thoroughly in [11] and [13]. In [11], two dimensionless parameters of the problem are discussed; i.e., the "Bueger number,"

$$N_{\mathrm{Bu}} \equiv \alpha L$$

and the "Boltzmann Number."

$$N_{\mathrm{Bo}} \equiv \frac{Cp\,\rho\,V}{\sigma T^3}.$$

These two parameters govern the combination of fluid convection and radiative transfer. Physically, these parameters have the following meaning:

$N_{\mathrm{Bo}} \to \infty$ implies a completely cold gas (i.e., $T \to 0$)

$N_{\mathrm{Bo}} \to 0$ implies a very hot gas

$N_{\mathrm{Bu}} \to \infty$ implies a completely opaque gas ($1/\alpha$ small)

$N_{\mathrm{Bu}} \to 0$ implies a completely transparent gas.

These upper and lower limits of the two parameters lead to limiting cases of sound wave propagation with radiative effects. If $N_{\mathrm{Bo}} \to \infty$ (completely cold gas) or $N_{\mathrm{Bu}} \to 0$ (completely transparent gas) the solution to the problem is the classical isentropic acoustic wave, because no radiative transfer takes place under these circumstances. If $N_{\mathrm{Bu}} \to \infty$ (completely opaque gas), the classical isentropic wave is again the solution, because in this case, although radiation may be intense, it is immediately reabsorbed near its point of emission, and once again no net radiative transfer takes place. If the gas is quite hot, and quite transparent, radiation tends to smooth out temperature gradients, and an acoustic wave travels at the isothermal sound speed rather than the isentropic sound speed. Thus, waves vary in their speeds of propagation, changing from the isentropic sound speed at small N_{Bu} to the isothermal sound speed in a range

near $1.5 N_{Bu} = N_{Bo}/(\gamma \cdot 8.33)$. Also, within this range of "velocity dispersion" the damping of the wave due to radiation reaches a maximum. For large values of N_{Bu}, the wave speed returns to the isentropic sound speed, and there may appear another sharp local peak in the damping.

In the analysis of harmonic waves given in [11], it appears that in addition to the classical acoustic wave traveling at either the isentropic or the isothermal sound speed, there is a second, radiation induced, harmonic wave, which has no counterpart in classical acoustic theory. The speed and damping of this wave are strongly dependent upon N_{Bo} and N_{Bu}. Its speed varies from infinite at $N_{Bu} = 0$, to zero at $N_{Bu} \to \infty$ for all values of N_{Bo}. The damping varies from zero at $N_{Bu} = 0$ to infinite at $N_{Bu} \to \infty$. For a fixed finite value of N_{Bu}, the damping goes from a finite value at $N_{Bo} \to \infty$ to zero at $N_{Bo} = 0$, at the same time the wave speed goes from a very high value to a low value, then back again to a very high value. In general, this second value has greater damping and higher speed than the classical wave, but at very high temperatures the damping may be comparable for both waves, and for a sufficiently opaque gas, the speeds may be essentially the same.

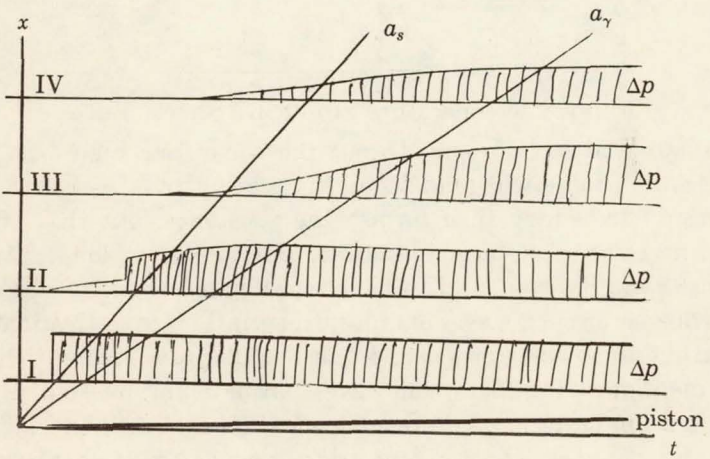


FIGURE 7. Pressures of a Jump Wave

In some ways, study of the progress of a single jump wave (rather than a harmonic wave) in a dispersive medium yields a plainer

picture of the trend of events. Figure 7 sketches Baldwin's result in [13] for the pressure of a jump wave produced by an impulsively-moved piston. Close to the piston (location I), the wave is only slightly dispersed, and $N_{Bu} \ll 1$, because the wave thickness (L) is small. Thermal energy will, however, begin to leak across the wave front, as indicated at location II, and a decay of the jump amplitude becomes evident. At III, where $L\alpha$ is no longer small, the front is very much flattened, and the wave progresses at the isothermal sound speed. At IV, the profile continues to spread out as we approach radiative equilibrium, but the wave center now travels at the isentropic sound speed again, and we have $q_R \to 0$, in the limit as $N_{Bu} \to 0$. A complete discussion of the foregoing problem is given in [13].

h. *Waves of Finite Strength.* Behind a strong shock wave there is a large, sudden temperature rise. The very hot gases behind the shock

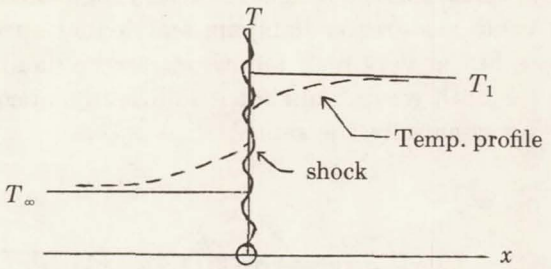


FIGURE 8. Temperature Profile in a Shock Wave

will radiate and tend to smooth out the wave (See Figure 8). The temperature profile will tend to be that shown by the dotted line. Radiation effects may then furnish the resistance, like that of viscosity, necessary for shock formation, but acting at longer range.

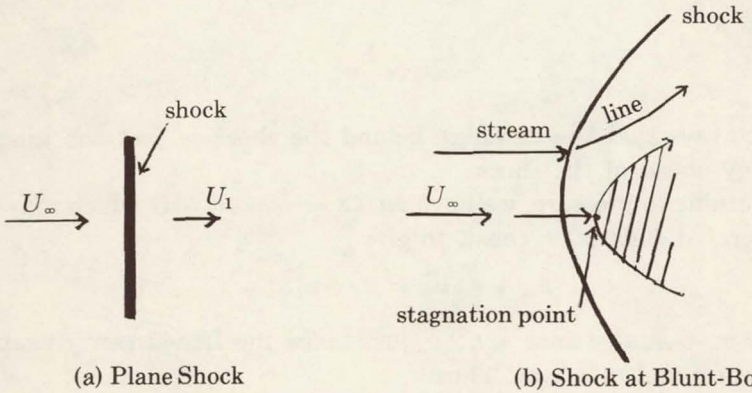
Heaslet and Baldwin, in [12], have investigated the effects of radiative transfer on such waves of finite strength. This investigation was carried out under the grey gas and quasi-equilibrium assumptions already mentioned. It is found that waves may be maintained entirely by radiation. The situation is quite analogous to heat addition in a constant area channel, wherein heat added always tends to drive the flow toward sonic speed. In this case, as in [6], both classical and radiation-induced waves are included, the entire spectrum of both being considered to make up the shock.

It is important to realize that radiative effects in flows can be

large, especially for superorbital speeds, and may play an important role. The analysis of these effects is very complicated, but possibly one may hope that some simplifying assumption will appear (similar to the Telegraph Equation assumption) in order to make these problems more tractable analytically.

III. **Plane shock waves.** We now turn to the problem of blunt body flows with shock waves and examine the effects of chemical nonequilibrium. We will also introduce models for the analysis of such flows.

a. *Description of Strong Shock.* First, we will briefly review shock waves and the Rankine-Hugoniot relations for very strong shocks. We distinguish between the two shock configurations shown in Figure 9(a) and (b). Figure 9(a) shows a freely propagating plane shock, such as in a shock tube. Figure 9(b) shows a bow shock about a blunt body traveling at hypersonic speed. In this case one has the additional complications of the turning of the flow around the body, which causes the shock to curve, and the existence of a stagnation



(a) Plane Shock (b) Shock at Blunt-Body

FIGURE 9. Two Shock Configurations

point. We shall deal with inviscid flows unless otherwise stated. We shall consider only the very strong shocks in which dissociative and radiative effects are important.

b. *Normal Shock Relations.* We now specialize to the case of a plane shock in one dimension. The differential equations (3), (6), and (7), neglecting transport effects, are

$$\text{Energy:} \quad \rho u \frac{\partial H}{\partial x} = u \frac{\partial p}{\partial x}$$

(25) $$\text{Momentum:} \quad \frac{\partial p}{\partial x} + \rho u \frac{\partial u}{\partial x} = 0$$

$$\text{Continuity:} \quad \frac{\partial (\rho u)}{\partial x} = 0.$$

The iterated forms of these equations apply across shock waves (we do not prove this here). Combining the energy equation with the momentum equation, we find

$$\rho \frac{\partial H}{\partial x} = - \rho u \frac{\partial u}{\partial x}.$$

Integrating, we find $H_\infty + \frac{1}{2}u_\infty^2 = H_1 + \frac{1}{2}u_1^2$, where ∞ signifies evaluation far ahead of, and 1 far behind, the shock. Now, in the free stream, $H_\infty \ll u_\infty^2$. Therefore, we may disregard H_∞. Similarly, $\frac{1}{2}u_1^2$ is small compared to H_1. Therefore, for very strong shocks we may write

(26) $$H_1 \cong \frac{1}{2}u_\infty^2$$

which says that the enthalpy behind the shock is just the kinetic energy ahead of the shock.

Turning to pressure, we have $\partial p/\partial x = -\rho u(\partial u/\partial x)$ which may be integrated since $\rho u = \text{const}$, to give

$$p_\infty + \rho_\infty u_\infty^2 = p_1 + \rho_1 u_1^2.$$

Now $p_\infty \ll \rho_\infty u_\infty^2$ (since $\rho_\infty u_\infty^2$ is just twice the free-stream dynamic pressure) and $p_1 \gg \rho_1 u_1^2$. Thus:

(27) $$p_1 \cong \rho_\infty u_\infty^2.$$

The pressure behind the strong shock comes from the conversion of essentially all the free stream momentum into a force, through deceleration of the flow.

We now assume an ideal gas for the purposes of an order-of-magnitude analysis, and note that $R/Cp = 1 - 1/\gamma$. For the temperature ratio, $H_1 = CpT_1 = \frac{1}{2}u_\infty^2$,

(28) $$\frac{T_1}{T_\infty} \cong \frac{\frac{1}{2}u_\infty^2}{CpT_\infty} = \frac{\gamma - 1}{2} M_\infty^2$$

(29) Pressure ratio: $$\frac{p_1}{p_\infty} \cong \frac{\rho_\infty u_\infty^2}{p_\infty} = \gamma M_\infty^2$$

(30) Density Ratio: $$\frac{\rho_1}{\rho_\infty} = \frac{p_1}{p_\infty} \cdot \frac{T_\infty}{T_1} \cong \frac{2\gamma}{\gamma - 1}$$

and we note that $2\gamma/(\gamma - 1)$ is approximately 10. It is important to note that the density ratio is finite, while the pressure and temperature ratios are unbounded, increasing as the square of the Mach number.

Now let $\frac{1}{2}(\gamma - 1) \equiv \epsilon$, a small quantity. We now investigate the order of magnitudes of changes in the variables behind the shock, since we are interested in the concentration of atoms, rates of pressure change, etc.

$$\frac{\partial}{\partial x}\left(\frac{p}{\rho_\infty u_\infty^2}\right) = \frac{u^2}{u_\infty^2}\frac{\partial}{\partial x}\left(\frac{\rho}{\rho_\infty}\right) \sim \epsilon^2 \cdot \frac{1}{\epsilon} = \epsilon.$$

Also, since $\rho_\infty u_\infty = \rho_1 u_1$, we have $u_1/u_\infty \sim \epsilon$. Thus, if ϵ is small, pressure varies more slowly than density. Also,

$$\frac{\rho}{\rho_\infty}\frac{\partial}{\partial x}\left(\frac{H}{u_\infty^2}\right) = \frac{\partial}{\partial x}\left(\frac{p}{\rho_\infty u_\infty^2}\right) \sim \epsilon$$

so

$$\frac{\partial}{\partial x}\left(\frac{H}{u_\infty^2}\right) \sim \epsilon \frac{1}{1/\epsilon} = \epsilon^2.$$

$p/(\rho_\infty u_\infty^2)$ is of order 1 behind the shock and the derivative with respect to the nondimensional distance x is of order ϵ. (Note that x is a distance characteristic of the relaxation thickness of the flow.) We have then, the following order of magnitude relations behind the shock:

$$\frac{p}{\rho_\infty u_\infty^2} \sim 1, \qquad \frac{\partial}{\partial x}\left(\frac{p}{\rho_\infty u_\infty^2}\right) \sim \epsilon,$$

$$\frac{H}{u_\infty^2} \sim 1, \qquad \frac{\partial}{\partial x}\left(\frac{H}{u_\infty^2}\right) \sim \epsilon^2.$$

These relations indicate that concentration and temperature "trade off" behind the shock, since ρ, c and T are the only quantities which vary appreciably. Enthalpy is very nearly constant, and pressure varies only slightly downstream of the shock.

We now examine the processes behind the shock, and inquire into the behavior of the concentration of atoms. We assume $T_D = 60{,}000°$ K for air. Now by Equation (10) for steady flow of a Lighthill gas, we have (following [**20**]):

$$(31) \qquad u\frac{\partial c}{\partial x} = \frac{W_1}{\rho} = \text{const}\, p\frac{T^{-(s+1)}}{1+c}\left[(1-c)e^{-T_D/T} - \frac{\rho}{\rho_D}c^2\right]$$

and we note that the square bracket goes to zero for equilibrium. This is the "production law" for atoms. Now when the gas is subjected to a step increase in temperature, by passing through a strong shock, the first chemical process will be dissociation, resulting from two-body (binary) collisions. Initially, recombination (a ternary process) will have little or no effect; however, as time passes recombination must become increasingly strong, and must finally balance dissociation at equilibrium.

c. *Binary Scaling.* For the present, we will confine our attention to that part of the flow immediately behind the shock where we can neglect recombination. Now, setting $W_1/\rho \equiv 1/\tau$, neglecting the second term in brackets, and noting that $p \propto \rho_\infty$, then we may write that $\tau_{\text{relax}} \propto 1/\rho_\infty$. Because $\tau_{\text{passage}} = L/u_\infty$, we have

$$\frac{\tau_{\text{relax}}}{\tau_{\text{passage}}} \propto \frac{u_\infty}{\rho_\infty L}.$$

This ratio is a parameter of the flow. $\rho_\infty L$ occurs in a group and constitutes a similarity parameter for Equation (31), recombination neglected. If we recall the simple case

$$c = c_{\text{eq}}(1 - e^{-t/\tau})$$

and replace t by x/u_∞ and $\tau \sim 1/\rho_\infty$, then we have

$$c = c_{\text{eq}}(1 - e^{-(x/L)\cdot(\rho_\infty L/u_\infty)\cdot\text{const}}).$$

If u_∞ is constant, then $\rho_\infty L$ becomes a scaling parameter for this relaxing flow. The same holds true for the more complicated production law given earlier, provided initial composition is fixed. We have what amounts to a restricted Reynolds similarity. This is not surprising, of course, because viscosity is a binary effect.

d. *Calculations for a Lighthill Gas.* Following Gibson in [**15**], we consider the case where the pressure and enthalpy are nearly constant. Now $u(\partial c/\partial x) = f(T)p[(1-c)/(1+c)]e^{-T_D/T}$ represents Equation (31) for steady flow and no recombination, and

$$T = \frac{H - h^{(0)}c}{R(4 + c)/m_2}$$

for a Lighthill gas. If H is constant (and it is, for all practical purposes, behind the shock) this means that $T = T(c)$, and

$$f(T) = F(c).$$

We now define a new variable χ such that

$$\chi \equiv \int_0^x \frac{p\,dx}{u}, \text{ so that } \frac{\partial}{\partial x} = \frac{p}{u}\frac{\partial}{\partial \chi}.$$

Substituting into the rate equation,

$$\frac{\partial c}{\partial \chi} = f(T)\frac{1 - c}{1 + c}e^{-T_D/T} \tag{32}$$

so that the right-hand side is a function of c only, and

$$\chi = I(c), \qquad \text{or} \qquad c = I^{-1}(\chi).$$

In this situation, if we plot the concentration c versus χ, we get (see [15] and [21]) the universal curve drawn by Gibson, as shown in Figure 10 (which is taken from [21]). Thus, a binary scaling scheme is found for strong shock waves. Note that the variable χ contains $\rho_\infty L$ as a parameter, and depends additionally on u_∞ and x/L.

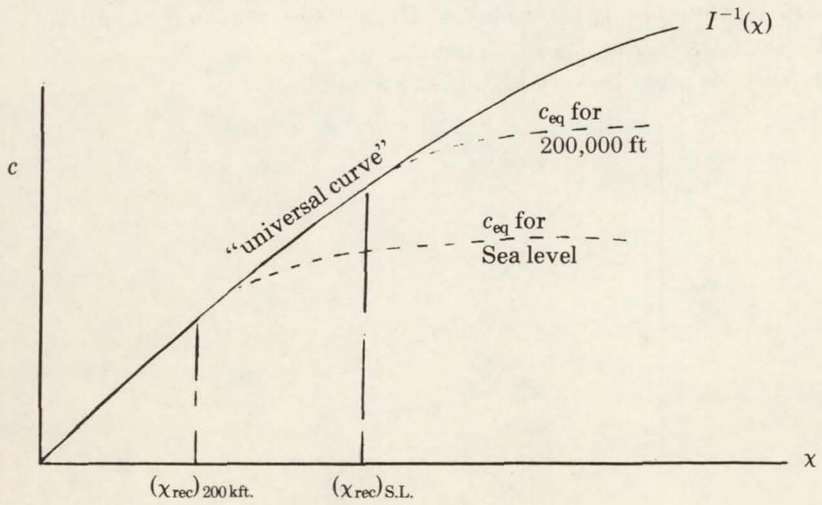


FIGURE 10. Gibson's Universal Curve

Gibson, in [21], was able to show that even the departures from the universal curve of Figure 10 can be predicted when binary scaling is not quite applicable. In the complete equation,

$$\frac{dc}{d\chi} = \frac{dc}{dI}\left(1 - \frac{\rho}{\rho_D}\frac{c^2}{1-c}e^{T_D/T}\right). \tag{32a}$$

The second term is replaced by I/I_∞ (I_∞ is the final equilibrium value of $I(c)$). Justification of this step is described in [21], for cases of high equilibrium dissociation level. The solution of Equation (32a) is, then,

$$I = I_\infty(1 - e^{-\chi/I_\infty}). \tag{33}$$

This remarkable formula recalls the simple relaxation law previously discussed, where $I(c)$ now plays the role of c. The effects due to recombination are shown as dotted lines in Figure 10, and appear only as departures from the universal plot near the end of the curve in question. Physically, we expect that the lower the density, the higher will be the proportion of binary collisions, i.e., those promoting dissociation. Recombination results from three-body collisions, and is proportional to ρ^3, and thus appears as a small effect near the ends of the curves. The scaling law renders the initial (dissociation) portion of the curves similar, regardless of initial density (altitude). For the scaling law to be useful, there are two required conditions:

1. Shock wave must be very strong.
2. Altitude must be very high (i.e., ρ small).

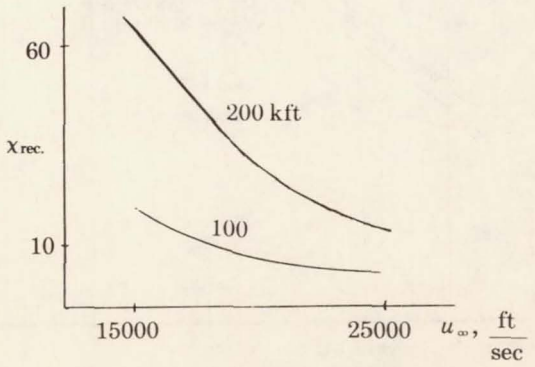


FIGURE 11. Scaling Limits

e. *Scaling Limit.* Clearly, as χ increases, the recombination effects become more important, and Figure 11 (taken from [21]) indicates qualitatively the limits of χ_{rec}, the value of χ beyond which the recombination term can no longer be neglected. The criterion chosen for Figure 11 is that the recombination rate is approximately 3/10 of the dissociation rate.

Referring to Equation (32a), we see that when ρ is small (high altitude) a larger value of c, and hence χ, can be reached before the second term is comparable to 1. Also, when the speed is lowered, the density is not greatly affected, but c_∞ is much less, so the effect of c^2 in Equation (32a) holds down the value of that term as χ increases, and again a higher value of χ is permitted.

f. *Calculations for Real Air.* The calculations just described for a Lighthill gas certainly emphasized the value of binary scaling for strong shock waves at low density. There have been many calculations of chemistry behind strong shocks in real air (see [16] and [17]). These may be presented in terms of the scaling variable χ. The set of reactions that have been studied are

$$
\begin{array}{lll}
O_2 \leftrightharpoons 2O & \left.\begin{array}{l}\\ \\ \end{array}\right\} & \text{Dissociative} \\
N_2 \leftrightharpoons 2N & & \\
N + O_2 \leftrightharpoons NO + O & \left.\begin{array}{l}\\ \\ \end{array}\right\} & \text{Exchange or} \\
NO + N \leftrightharpoons N_2 + O & & \text{"Shuffle" reactions} \\
N + O \leftrightharpoons NO^+ + e^- & & \text{Ionization.}
\end{array}
\tag{34}
$$

In general the "shuffle reactions" are predominantly to the right, which promotes the formation of oxygen atoms, and encourages the depletion of N by recombination into N_2 or NO. In other words, the the mechanism for the recombination of nitrogen is more powerful than the recombination mechanism for oxygen. The shuffle reactions are binary in nature, and thus have particular importance at high altitude.

Figure 12 (taken from [21]) shows various concentrations plotted versus χ, as is done in [21]. These are simply sketches, and details should be sought in [21].

The sketches indicate that binary scaling works for real air about as well as for the Lighthill gas. Also illustrated in Figure 13 taken from [21]), is the phenomenon of overshoot of NO and e^-, typical of high-energy flows (here, the shock speed was 23,000 fps).

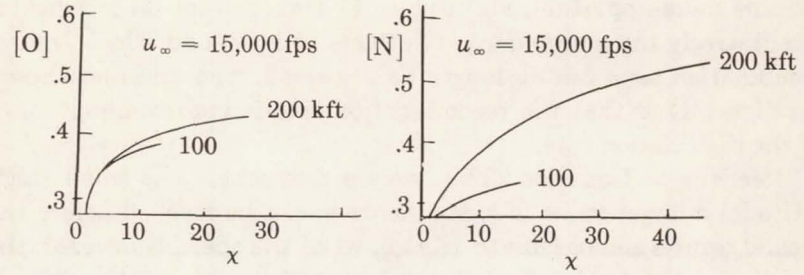


FIGURE 12. Concentrations

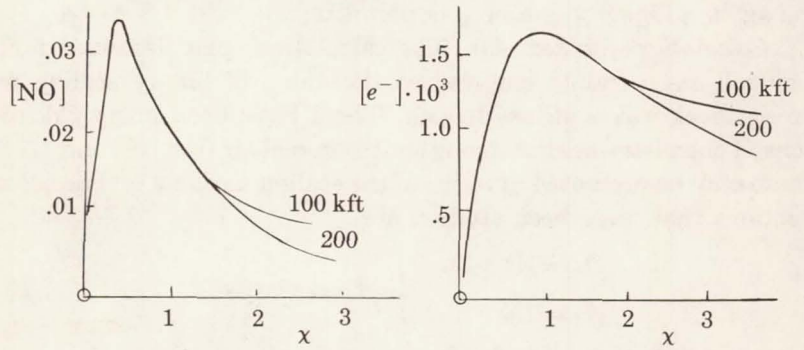


FIGURE 13. Overshoot

The overshoot of electron concentration may be understood as a consequence of the abnormally high translational temperature just behind the shock. Since translation equilibrates first, the ideal-gas temperature is reached-about 25,000°K for 23,000 fps. at 200,000 ft. altitude. In about 5mm behind the shock, O dissociation manages to take up its equilibrium share of energy, and, by then, the translational temperature has dropped to 8000°K. Of course, while the temperature is at the higher level, radiation and ionization are intense.

g. *Vibrational Coupling.* Actually, the overall process of equilibration is strongly affected by vibration. Figure 14 sketches the sequence: First vibrational excitation begins to rise (rotation is nearly as fast as translation), and dissociation follows. But, since dissociation occurs chiefly from excited vibrational states, vibration and dissociation are kinetically coupled (see [18]). Also, when dissociation does occur, it depletes the higher vibrational states (see [19]). Thus, at high energy and low density, the kinetic model

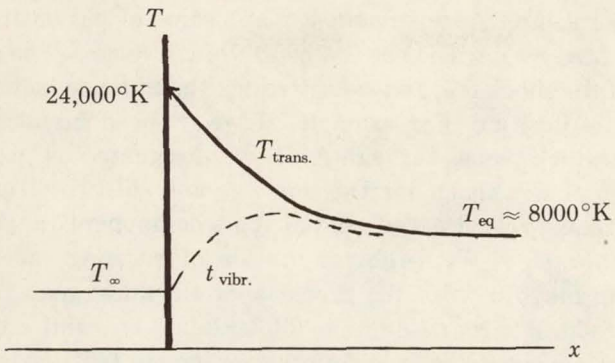


FIGURE 14. Vibrational Effects

for air must include vibration. We note, however, that vibrational excitation is a binary process, and the previous scaling considerations apply.

IV. **Blunt-body flows.** We turn now to problems of the inviscid flow about blunt bodies moving at hypersonic velocity, and consider the effect of chemical reactions. We shall be concerned with very strong shocks, as before, but now the shock is curved about the body, as sketcked in Figure 15.

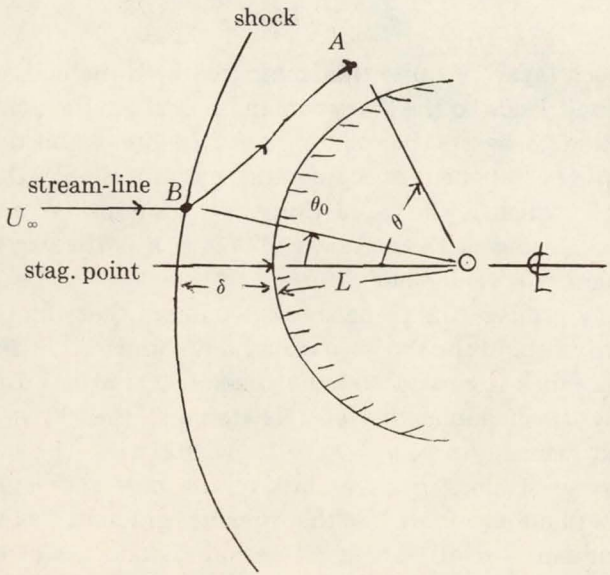


FIGURE 15. Blunt-body Shock Layer

a. *The Newtonian Approximation.* The radius of curvature of the nose is denoted by L with O as the center of curvature. The standoff distance of the shock is δ, and we introduce the independent variable θ to measure position. For example, if we wish to investigate the streamline which passes through A, B, we designate the position of penetration of the shock by the angle θ_0 and later position by θ. The free stream velocity is u_∞ and at A its component tangential to the body is $u_\infty \sin\theta$. With the assumption of a strong shock wave, the order of magnitude of the shock layer thickness may be determined as follows: From Equation (30), $\rho/\rho_\infty \sim 1/\epsilon$, and ϵ is rather small. Now, consider a cylinder in the free stream (with axis parallel to the free stream velocity), subtended by an angle θ. The mass flow through a cross section of this cylinder is approximately

$$\rho_\infty u_\infty (L \sin\theta)^2 \pi.$$

The tangential flow in the region between the shock and the body, at a location θ, is approximately

$$\rho u_\infty \sin\theta (\pi L \sin\theta)\delta.$$

These flows must be equal, and after cancellation, we have

$$\frac{\delta}{L} = \frac{\rho_\infty}{\rho} \sim \epsilon$$

and the "shock layer" is quite thin compared with the body dimension. This result leads to the "Newtonian" model for the analysis of blunt-body flows, wherein the colliding particles are assumed to give up their normal component of momentum on impact with the body (or the shock, which is the same thing, by assumption) and subsequently move tangent to the surface. This is a particular kind of reflection, neither specular nor diffuse.

For a freely propagating plane shock, we found that the pressure changes slowly behind the shock. With a body immersed in the flow, however, the shock is curved and the pressure variations are large. We have, by direct application of "Newtonian" theory, $p/(\rho_\infty u_\infty^2) \propto \cos^2\theta$, and hence $\partial(p/\rho_\infty u_\infty^2)/\partial x \sim 1$, where x is the ratio of distance measured along a streamline to the nose radius L. Previously, for a plane shock, we had that pressure gradient was of order ϵ (x being measured relative to a relaxation distance). Correspondingly, we now have $\partial(H/u_\infty^2)/\partial x \sim \epsilon$ (as against ϵ^2 for the plane

shock), and, from the momentum equation, $u/u_\infty \sim \sqrt{\epsilon}$ (as against ϵ for the plane shock).

The Newtonian theory requires modification to account for centrifugal forces and effects of shock-layer thickness. Centrifugal force corrections for a sphere (see [20]) modify the result $p \propto \cos^2\theta$ so that $p = 0$ (a sort of centrifugal separation) is predicted to occur at $\theta = \pi/3$.

b. *Binary Scaling.* If we re-examine Equation (32), we see that the crucial thing for binary scaling is that H remain nearly constant. This is so for blunt-body flows, because $\gamma - 1$ is small. The pressure is not constant, but its variation can be absorbed in χ. Thus, Gibson and Marrone in [21] show that we can carry over the binary scaling analysis to blunt-body cases, c being given by the same function of χ:

$$c = I^{-1}(\chi),$$

where

$$\chi \equiv \int \frac{pds}{u}, \tag{34a}$$

s being measured along a streamline. Finding p and u from Newtonian theory, we observe the binary scaling rule that results:

$$\chi = \int \frac{pds}{u} = \rho_\infty L \, \mathrm{fn}(u_\infty, \theta).$$

c. *Scaling Limits.* From Equation (34a), we see that as we approach the stagnation point of the body, $u \to 0$ and thus $\chi \to \infty$, so we cannot expect binary scaling to apply right there. On the other hand, as we travel along a streamline like $B - A$ in Figure 15, we know that $p \to 0$ at 60° (for a sphere), and thus χ reaches some limiting value, which might be less than χ_{rec}. Figure 16, which sketches information in [21], compares $\chi_{\max}$ and χ_{rec} under a number of conditions. For $L = 30$cm and altitude 200 kft, the limiting value of χ along the streamline $\theta_0 = \arcsin(1/5)$ just equals χ_{rec}, the scaling limit for that altitude. Thus, for streamlines beginning farther from the axis, the universal curve of Figure 10 applies all the way to 60°, while along streamlines closer to the axis the function $c(\chi)$ will break away from the universal curve.

Of course, even right at the axis ($\theta_0 = 0$), the actual departures

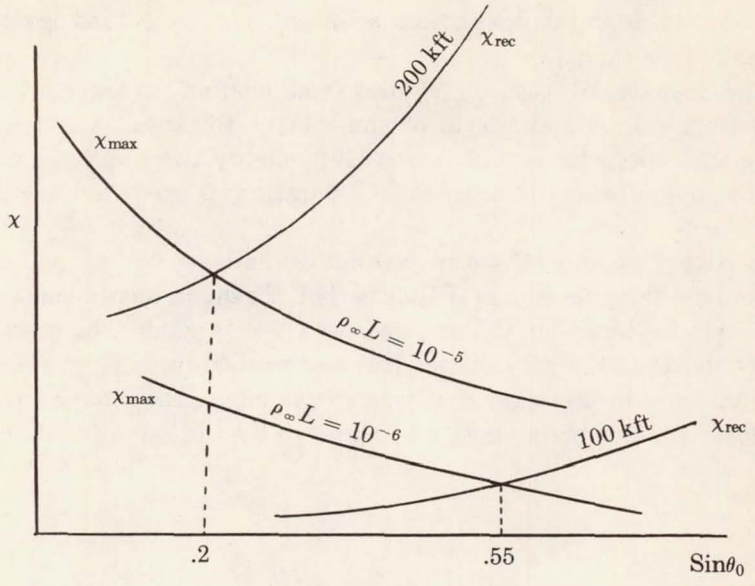


FIGURE 16. Limits on χ

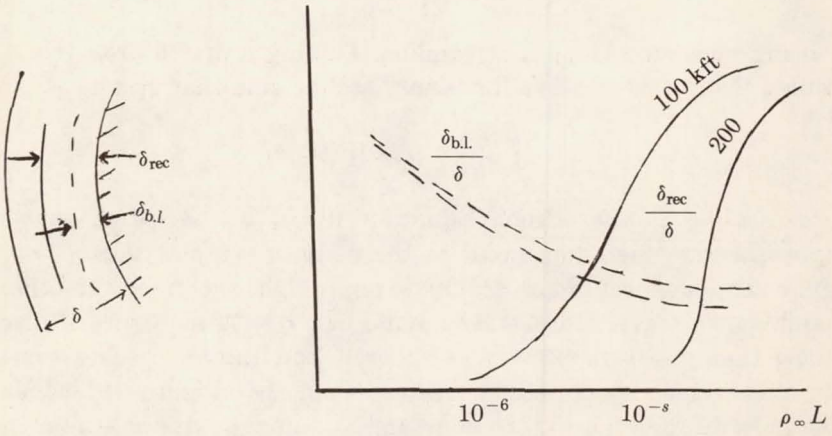


FIGURE 17. Recombination and Boundary Layer Thicknesses

from binary scaling occur only near the surface, so one may speak of a recombination thickness δ_{rec} which is perhaps a small fraction of the shock-layer thickness, or "stand-off distance" δ. Figure 17 (taken from [21]) shows how δ_{rec}/δ varies with $\rho_\infty L$. It appears that recombination plays a negligible role at 200 kft, if L is less than that

for which $\rho_\infty L \approx 10^{-5}$. Viscous effects at the nose should be included. They scale completely with $\rho_\infty L$, as indicated by the boundary layer thickness $\delta_{b.l.}$ in Figure 17.

In summary, we can use binary scaling for the nose region, at high altitude ($\rho_\infty L$ small) for flow suddenly heated by a strong (hypersonic) bow shock, concluding that dimensionless variables depend on $\rho_\infty L$ for the same u_∞ and same initial composition (see [21]). Under these conditions nonequilibrium chemistry and ionization may be analyzed quite simply.

d. *Example: Stand-off Distance.* Figure 18 (taken from [21]) shows Newtonian calculations of δ/L showing the expected scaling for small values of $\rho_\infty L$. Also shown are two points gotten from computed solutions with full air chemistry. Their agreement with the scaling rule, even for scales differing by 100, is an excellent indication of general applicability of the binary model.

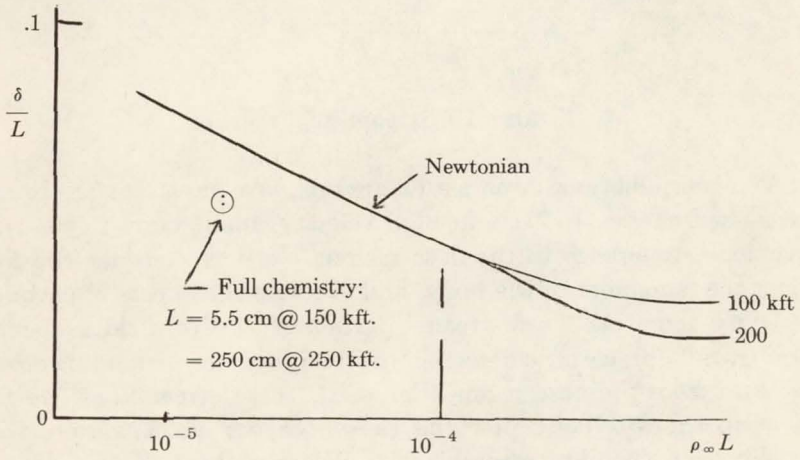


FIGURE 18. Stand-off Distance

e. *Exact Solutions for Blunt Bodies.* The "exact" calculations just mentioned were done by a method due to Lick (see [22]), whereby a shock shape is assumed, and the full equations with chemistry are numerically integrated inward to "find the body." Hall, Eschenroeder, and Marrone (see [17]) made calculations by this method, using the 5 reactions cited previously. Here, we encounter a feature not present in the plane shock problem: Now, as the flow proceeds, it cools, and the recombination may quench before equilibrium is

reached. Figure 19 (taken from [17]) shows this effect for two streamlines, at $u_\infty = 23{,}000$ fps. The "faster" streamline, B, shows more freezing effect than A, as we would expect. The freezing of N is more pronounced because the exchange reactions tend to deplete N, and keep O high. Thus reactions are binary, and are the favored recombination reactions at low density.

The freezing effect just described is important to assess in preparation for analyzing afterbody and wake flows.

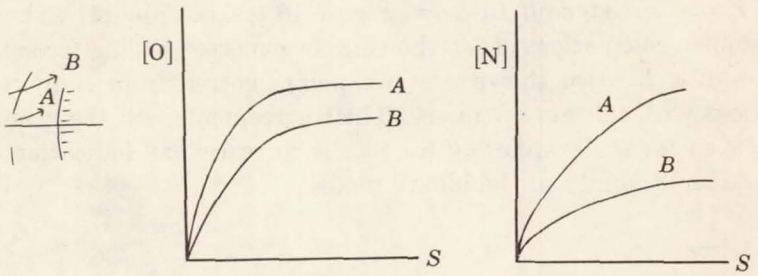
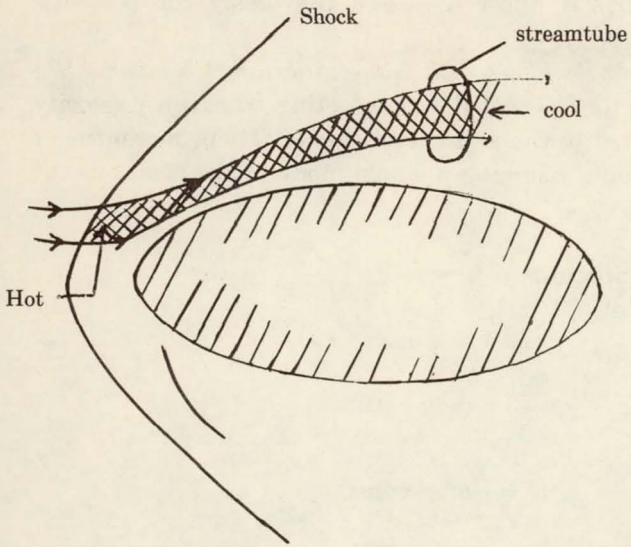
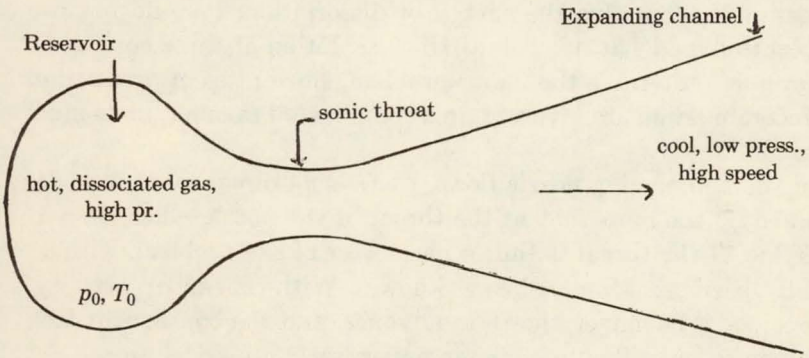


FIGURE 19. Streamline Freezing

V. **Nonequilibrium in nozzle expansions.** We have, so far, briefly discussed chemical effects in high velocity, blunt-body flows, with particular reference to the nose region. We now consider the flow over the remainder of the body, and examine the effects of chemical activity upon this "downstream" portion of the flow field. Referring to Figure 20(a), we consider a pair of stream surfaces extending downstream from the nose region. Following these streamlines, we find that in the afterbody flow the gas undergoes an expansion, with cooling and chemical recombination. We can think of this flow as being like the expansion of a very hot gas through a nozzle (see [23] and [24]), the nose region corresponding to a reservoir of hot dissociated gas. We, then, learn much about the flow over afterbodies by considering channel flows, as illustrated in Figure 20(b). Of course, such channel flow studies are of interest not only for hypersonic applications, but also for rocket nozzles and shock tunnels.



(a) Flow between Streamlines



(b) Flow through a Nozzle

FIGURE 20. Analogous Flows

Expansion and cooling experienced by the gas in passing through a nozzle of course results in chemical recombination and reduction of the high levels of other modes of excitation. Thus, processes occur

which are the reverse of those discussed previously for the nose region of the blunt body.

a. *Equation for Nozzle Flows with Nonequilibrium Chemistry.* We write the equations for channel flow, neglecting transverse velocity components compared to the axial components. Then, measuring x down the channel, and assuming a single mode of dissociation, the equations reduce to

(35)
$$\text{Continuity of Species:} \quad u\frac{\partial c}{\partial x} = \frac{W_1}{\rho}$$

$$\text{Continuity:} \quad \rho u A = \text{const}$$

$$\text{Momentum:} \quad \frac{dp}{\rho} + u\,du = 0$$

$$\text{Energy:} \quad H = \frac{1}{2}u^2 = \text{const}$$

$$\text{State:} \quad p = (1 + c)\rho RT.$$

We consider, as usual, that the enthalpy H consists of the internal degrees of freedom plus the energy of dissociation. (We do not use the Lighthill model at this point.) In the solution of these equations, the production term is the major problem, since rates of production and recombination are involved, in a complicated manner, in channel flows.

For the nonreacting nozzle flows, there is a critical mass flow determined by the mass flow at the throat of the nozzle when $M = 1$. Mass flow at the throat is thus a parameter of the problem, with a^* (sound speed at $M = 1$) being known. With chemical activity, however, a^* is no longer known in advance, and the consequent lack of a definite mass-flow parameter makes real numerical (machine) solutions for the problem, via forward integration from equilibrium, very difficult. Another difficulty concerns initial departures from equilibrium, where $\partial^n c/\partial x^n = 0$. Computationally, one has something similar to an essential singularity; thus, integration away from equilibrium tends to require very fine-grained calculations.

Early work in the field of nonequilibrium channel flows was done by Bray in [**23**], and by Hall and Russo in [**24**]. Hall and Russo used, first a Taylor series expansion of about 10 terms in $1/\sqrt{A}$ (A is area $\div$ throat area) to carry the calculations to the throat. By iteration of

these results, a mass flow was established. Then, downstream of the throat, a modified Runge-Kutta scheme was used. For a hyperbolic channel shape,

$$A = 1 + \frac{x^2}{L^2},$$

where L is a suitable length parameter. The results from [24] appear in Figure 21 as follows.

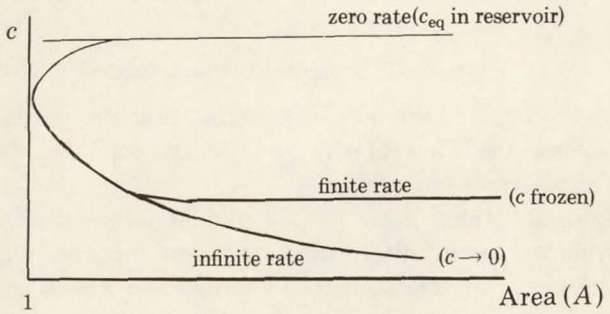


FIGURE 21. Atom Concentration against Area Ratio

If equilibrium is maintained throughout the flow, which implies a recombination rate approaching infinity, we get the "equilibrium" curve called "infinite rate" in Figure 21. As a result of cooling and expansion, density decreases; we recall from Equation (31) that

$$\frac{W_1}{\rho} \propto \left[(1 - c)e^{-T_D/T} - \frac{\rho}{\rho_D} c^2 \right].$$

Since density decreases, the rate of recombination decreases, therefore there is a point where the process of recombination cannot keep up. (That is, the three-body collisions necessary for recombination become too infrequent to maintain equilibrium.) The result is that freezing occurs, rather suddenly, as shown in the curve called "infinite rate" in Figure 21. Concentration is constant downstream of the freezing point. Such freezing of the flow is of great importance in two cases:

1. *Propulsion systems.* In a rocket nozzle, the freezing results in the loss of kinetic energy, because some of the energy is frozen in the form of dissociation and vibration energy, and is thus not available for propulsion.

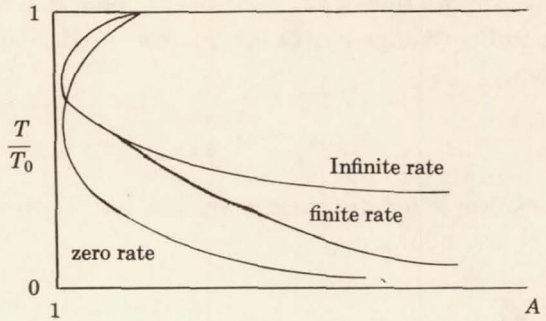


FIGURE 22. Temperature against Area Ratio

2. *Hypersonic wind tunnels.* In such test tunnels, we do not wish to have a dissociated flow at the test section, and therefore we must care to avoid too-early freezing.

In Figure 22 (taken from [24]) a plot of temperature versus area ratio is shown for infinite, finite, and zero rates of recombination. (T_0 is the reservoir temperature.) Comparison with Figure 21 shows how temperature and concentration "trade off."

A plot of pressure versus area ratio is shown in Figure 23 (taken from [24]).

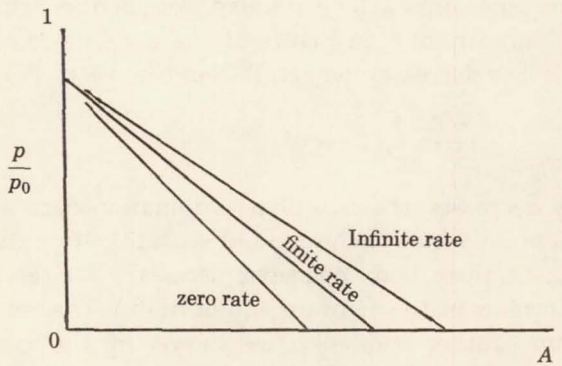


FIGURE 23. Pressure against Area Ratio

We note that the difference between infinite and zero rates is not as great as for temperature. Thus, velocity is not coupled to concentration as closely as is temperature. The foregoing figures are based on a temperature of 6000° K and a pressure of 100 atmospheres in the reservoir.

The final frozen level of concentration is lower for longer channels of the same shape, since the longer the channel, the greater the time during which the recombination process can proceed. Plotting length versus frozen concentration, we have a relationship as shown in Figure 24 (taken from [24]).

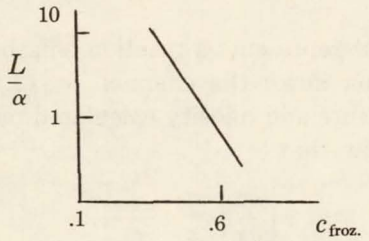


FIGURE 24. Length against Frozen Concentration

b. *Approximate Freezing Criteria.* It is of interest to predict the occurrence of freezing in the channel, but since channel flows are characterized by recombination processes, there is no possibility of applying a binary scaling law.

The change of concentration along an equilibrium path is fixed by the geometry of the channel. We must now inquire whether the recombination rate will be sufficient to maintain equilibrium, and if not where along the channel freezing will occur. This problem has been studied by Bray in [25], and Hall and Russo in [24]. Bray assumed that the rate for equilibrium flow is given in terms of channel geometry, etc., so that $(Dc/Dt)_{eq}$ is known. One then asks when this is greater than the recombination rate:

$$\left|\frac{Dc}{Dt}\right|_{eq} > \text{const}(\rho^2 T^{-s} c^2)_{eq},$$

where $\rho^2 T^{-s} c^2$ is the recombination rate term. The left-hand side is an aerodynamic requirement, and we ask when it is greater than a quantity proportional to $\rho^2 T^{-s} c^2$. Bray employed the equilibrium mass action law for a Lighthill gas to eliminate the density, so that

$$(\rho^2 T^{-s} c^2)_{eq} \sim \left(\frac{1-c}{c}\right)^2_{eq} e^{-2T_D/T} T^{-s}.$$

Then Bray's freezing criterion is

$$\left(\frac{1-c}{c}\right)^2 \frac{e^{-2T_D/T}}{(1+c)T} = \text{const.} \tag{36}$$

Hall and Russo took the freezing criterion to be

$$\frac{\partial c}{\partial x} = -\frac{c - c_{\text{eq,finite}}}{r}, \tag{37}$$

where r (a function) represents a reaction length or typical path for relaxation increasing down the channel. $c_{\text{eq,finite}}$ is the (fictitious) c_{eq} for the temperature and density calculated on a finite-rate basis. One may easily show that

$$\left|\frac{\partial c}{\partial x}\right| \begin{matrix} \ll \\ = \\ \gg \end{matrix} \frac{c_{\text{eq,finite}}}{r},$$

where the upper inequality holds for equilibrium flow, and the lower inequality holds for frozen flow. The freezing criterion is defined, arbitrarily, by the equal sign, giving Equation (37). Reference to Figure 25 indicates these three cases. Figure 25 is taken from [**24**].

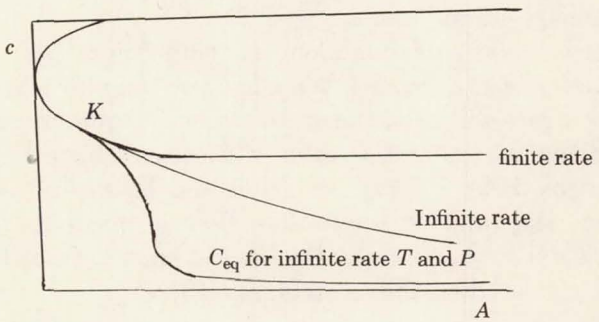


FIGURE 25. Concentration against Area Ratio

At point K (Figure 25) where the curve splits into three possible paths, we can predict $\partial c/\partial x$. The average rate during the abrupt freezing process is $\frac{1}{2}(\partial c/\partial x)_{\text{inf}}$. Also since $c_{\text{eq,finite}}$ drops so quickly to zero, its average value is $\frac{1}{2}c_{\text{inf}}$. The freezing criterion is then

$$\left(\frac{\partial c}{\partial x}\right)_{\text{inf}} = \left(\frac{c}{r}\right)_{\text{inf}}.$$

Machine calculations show the freezing process to be rather sudden, and at point K the freezing line breaks sharply away from the infinite or finite rate equilibrium lines.

c. *Entropy and the Sudden-Freezing Model.* The question of the entropy of the flow, while one of great interest, is also one of considerable difficulty. Entropy is defined only for a state which can be reached by an equilibrium process. For the early, equilibrium flow the entropy is well defined, and the expansion is isentropic. The change from equilibrium to frozen flow is one which is nonisentropic, unless the freezing is mathematically sudden. However, completely frozen flow is again isentropic, because chemistry no longer participates.

The formula for entropy of a gas in equilibrium is

$$dS = \frac{dE + p\,dV}{T}. \tag{38}$$

Using Lighthill's gas and the state equation, we get

$$\frac{dS}{R} = 3\frac{dT}{T} + T_D\frac{dc}{T} - (1+c)\frac{d\rho}{\rho}. \tag{38a}$$

This form, as it stands, cannot be integrated to give a variable of state, except for frozen concentration, that is, for c constant. Then,

$$\frac{S_{fr}}{R} = 3\ln T - (1-c)\ln\rho + \text{const.} \tag{39}$$

For equilibrium, we have another relation,

$$\frac{c^2}{1-c} = \frac{\rho_D}{\rho}e^{-T_D/T}$$

by which one of the original three variables may be eliminated. Thus,

$$\frac{S_{eq}}{R} = 3\ln T + (1+c)\frac{T_D}{T} + c + 2\ln\frac{c}{1-c} + \text{const} \tag{40}$$

applies for equilibrium. The constant in Equation (39) may be chosen so that S_{eq} goes over to S_{fr} continuously at a sudden freezing point.

d. *The Mollier Diagram.* Bray noted in [**25**] that his freezing criterion, Equation (36) gives another relation between c or T, which, substituted into Equation (40), gives the freezing state in terms of S only:

$$c_{fr} = c(S_0); \quad T_{fr} = T(S_0). \tag{41}$$

Thus, one may construct a Mollier-type diagram, as shown schem-

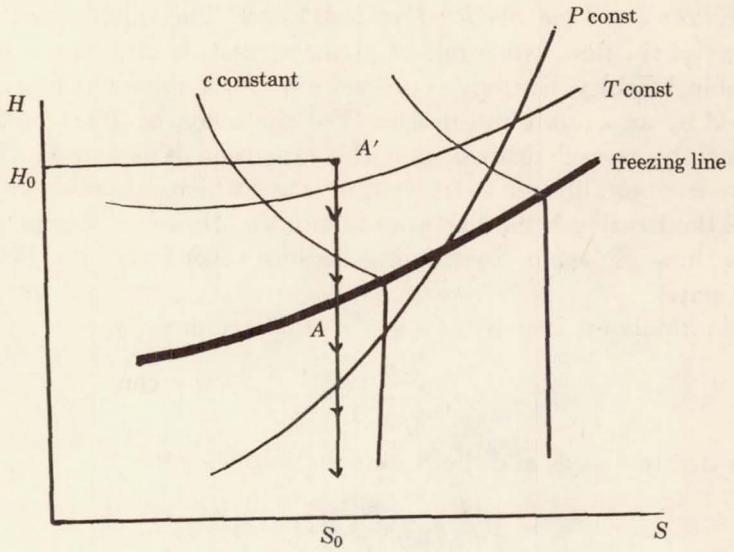


FIGURE 26. Mollier Diagram

atically in Figure 26 (taken from [25]).

According to this diagram, if we start a flow from a point of known enthalpy, A', the process proceeds along the vertical isentrope through the "freezing line." Actually, of course there may be a narrow freezing zone rather than a freezing line. In any case, this kind of diagram represents a very useful condensation of results, when freezing is sudden.

An approach to the problem of more gradual freezing has been made in [26] by considering two subsystems which flow out of equilibrium with each other. Then one can define entropy for these subsystems, each of which is internally in equilibrium. However, this is a dubious approach.

e. *Calculations for Real Air.* More complete machine calculations have been made in [34], including the five reactions (Equation (34)) important at high temperature. As usual, the exchange or shuffle reactions favor recombination of N. Results of calculations for the concentrations of the various species are shown in Figure 27 (taken from [34]). An important feature of such a flow is that the recombination of nitrogen is vigorous, via the NO reactions, and O concentration, as a result, stays high, freezing almost immediately.

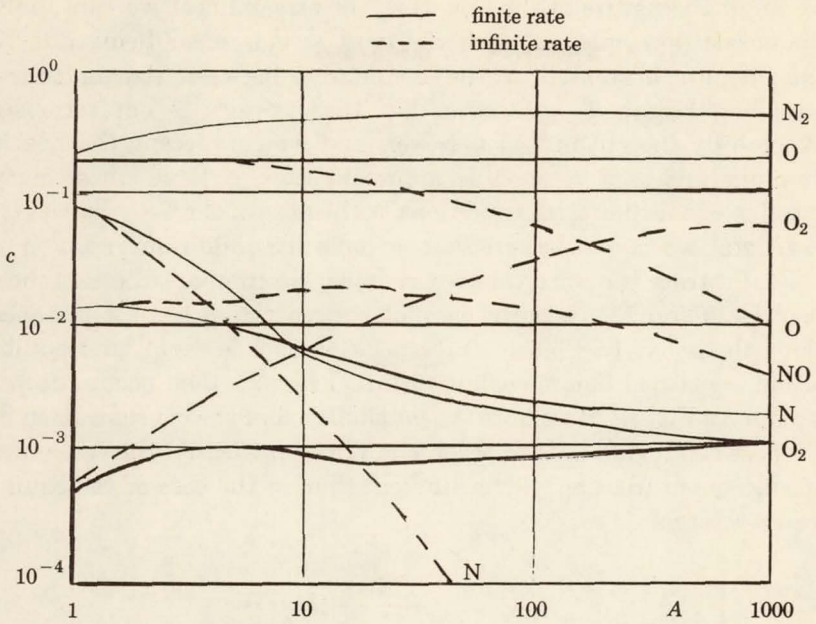


FIGURE 27. Concentration against Area Ratio for Air

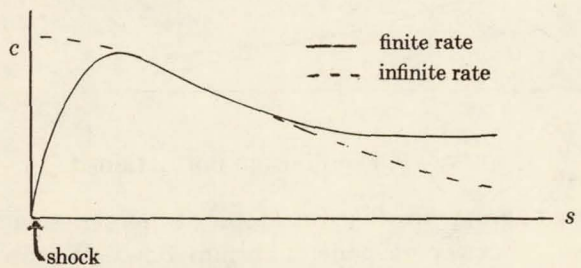


FIGURE 28. Equilibration and Subsequent Freezing

f. *Stream Tube Applications.* We now attempt to apply the nozzle results to flows about bodies. Referring to Figure 20(a), we return to the idea that the flow through a stream tube about the body is essentially described by a nozzle flow model such as we have just been discussing. The region of hot dissociated gas near the stagnation point serves as a high pressure, high energy reservoir. In general, the nozzle flow analysis cannot be applied without some reservations, due to the transverse gradients existing in the flow field. If, however,

we know the nature of the flow (i.e., the streamlines) we can apply the nozzle flow analysis, with the most serious error being due to the pressure mismatch at the boundaries between stream tubes. Recalling Figure 23, we note that the pressure is not seriously affected by the chemistry, however, and we can accept the nozzle flow analysis as a reasonable approximation. A large uncertainty exists in connection with conditions at the nose of the body, however; in general, we cannot be sure that we have an equilibrium reservoir.

To illustrate the uncertainties involved we compare the sketches, Figures 28 and 29. In both, we plot concentration against distance along the body. In Figure 28 dissociation rate is rapid, and equilibrium is reached before cooling begins. Freezing then occurs downstream. In Figure 29 we note the possibility that the concentration in the nose cap region fails to reach equilibrium. Clearly, the dynamics of subsequent freezing will be different than in the case of the equilibrium reservoir.

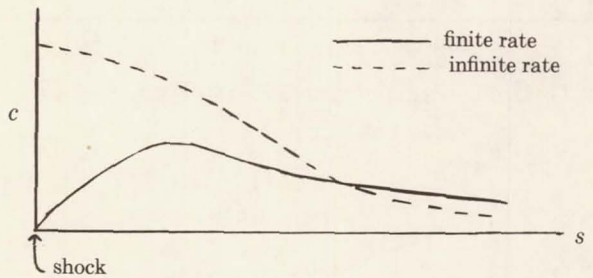


FIGURE 29. Equilibrium not Attained

VI. **Transport properties.** The final topic of this series will deal with the effects of viscosity on nonequilibrium flows. We first consider transport properties, and in §VII will deal with viscous flows. We will attempt to find flow models for nonequilibrium flows including viscous terms. We have already noted that chemical activity and radiative transfer result in phenomena not unlike viscous effects, i.e., damping and dispersion in acoustic waves and shocks. The time lags encountered in chemical kinetics relate to the time lags in viscous effects.

a. *Equations of a Binary Mixture.* We shall consider a multicomponent high temperature gas (in particular, air). The transport quantities appear as terms on the right-hand side of Equations (4),

(6), and (7). Now these are rewritten as

$$-\frac{\partial}{\partial n}(\rho c u_n) = \frac{\partial}{\partial n}\left(\rho D \frac{\partial c}{\partial n}\right)$$

$$\frac{\partial}{\partial n}\left(\mu \frac{\partial u}{\partial n}\right) \tag{42}$$

$$\mu\left(\frac{\partial u}{\partial n}\right)^2 + \frac{\partial}{\partial n}\left[\lambda \frac{\partial T}{\partial n} + \rho D\left(\frac{\partial c}{\partial n}\frac{H_1}{c_1} - \frac{\partial c}{\partial n}\frac{H_2}{1-c_1}\right)\right].$$

We indicate the physical meaning of these terms:

$\frac{\partial}{\partial n}\left(\rho D \frac{\partial c}{\partial n}\right)$ is a concentration gradient effect;

$\frac{\partial}{\partial n}\left(\mu \frac{\partial u}{\partial n}\right)$ is the viscous shear stress term; and

$$\mu\left(\frac{\partial u}{\partial n}\right)^2 + \frac{\partial}{\partial n}\left[\lambda \frac{\partial T}{\partial n} + \rho D\left(\frac{\partial c}{\partial n}\frac{H_1}{c_1} - \frac{\partial c}{\partial n}\frac{H_2}{1-c_1}\right)\right]$$

consists of a term to account for enthalpy increase due to dissipation, a heat conduction term, and a term to account for diffusion of enthalpy.

Enthalpy flux arises from the unequal transport of atoms and molecules. If more atoms than molecules are transported across a surface, then, since the atoms carry the dissociation energy, $h^{(0)}$, there is a net flux of enthalpy across the surface. Thus the final term consists of the diffusion velocity quantity, $\rho D(\partial c/\partial n)$, and $h^{(0)}$ is the property transported. We assume that $h^{(0)}$ is much greater than the internal energy of molecules, i.e., $H_{\text{int}} \ll ch^{(0)}$.

From the Equation (42) we see that we must specify three mixing parameters, D, λ, and μ. We now define two important dimensionless parameters:

Prandtl number:

$$\text{Pr} \equiv \frac{\mu Cp}{\lambda} \tag{43}$$

which compares viscous effects to heat transfer effects; and

Lewis number:

$$\text{Le} \equiv \frac{\rho Cp D}{\lambda} \tag{44}$$

which compares diffusion effects to heat conduction. The Prandtl number and Lewis number are both of unit order, and $\Pr \approx 3/4$. The Lewis number is frequently approximated by unity; however; this may not be very accurate, since the Lewis number depends upon temperature approximately as shown in Figure 30.

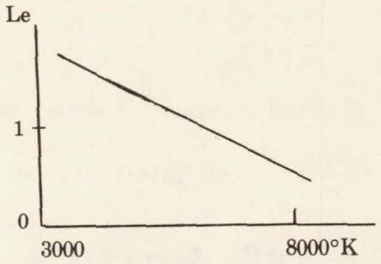


FIGURE 30. Variation of the Lewis Number

Using the nondimensional parameters, we note that

$$\frac{\partial}{\partial n}\left(\rho D \frac{\partial c}{\partial n}\right) = \frac{\partial}{\partial n}\left(\frac{\mathrm{Le}}{\Pr}\mu \frac{\partial c}{\partial n}\right) \tag{45}$$

and

$$\mu\left(\frac{\partial u}{\partial n}\right)^2 + \cdots = \mu\left(\frac{\partial u}{\partial n}\right)^2 + \frac{\partial}{\partial n}\left\{\frac{\mu}{\Pr}\left[\frac{\partial H}{\partial n} + (\mathrm{Le} - 1)h^{(0)}\frac{\partial c}{\partial n}\right]\right\}$$

noting that $CpT = H - h^{(0)}c$.

If, as in [27] we assume $\mathrm{Le} = 1$, the last term on the right-hand side vanishes, and the energy equation is much simplified, since the heat flux and energy equation do not explicitly involve chemistry. We assume, then, that diffusion and viscosity effects are so related that $\mathrm{Le} = 1$.

We note at this point that gas viscosity increases with an increase in temperature, as opposed to the decrease in viscosity with temperature in liquids. One may explain this behavior by considering the interpenetration of particles, which is greater for more energetic (higher temperature) gas particles. Thus a hot, or rarefied gas has a high viscosity. In dissociating flows, one finds that atoms have a greater penetrating depth than molecules, and therefore the dissociated gas is more viscous than undissociated gas. This is shown qualitatively in Figure 31, which is taken from [38].

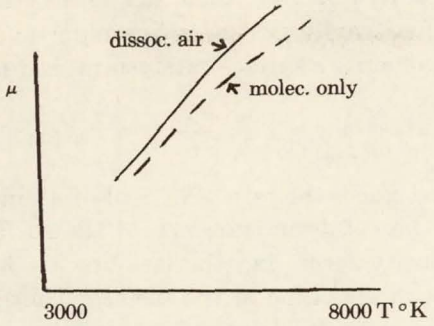


FIGURE 31. Variations of Viscosity

The collisional models used to calculate the viscosity of multicomponent, high-temperature gases are various; the matter is reviewed in [37]. Actually, it is usually more important to have the right model for high-temperature molecular collisions than it is to account for the presence of atoms (see [38]). For flow problems, viscosity enters through the product $\rho\mu$, and ρ is a stronger function of c (through the state equation) than is μ. In any case, a simple perturbation formula (see [38]) accounts quite well for the dependence of μ on c:

$$\mu = \mu_{\text{molec. only}}[1 + 0.3c + O(c^2)].$$

An important question at this point is: Where do we measure viscosity in flows with large temperature variations? We wish to use a constant value for viscosity (or, rather density times viscosity), but what value is most representative? In general, the answer is to evaluate viscosity in the hottest part of the flow. This is because the hottest part is also (usually) the region of lowest density, and hence forms a very thick layer compared with cooler parts. Properties of the thick layer dominate transport, and the viscosity of this layer rather well represents the effective viscosity of the whole flow.

b. *Surface Catalysis.* Consider a semi-infinite region bounded by a solid surface, and denote the surface conditions by the subscript zero. If an atom strikes and adheres to the surface, a second atom may strike it and recombine the two then leaving the surface as a molecule. In such a recombination, a three-body collision is not required, since the wall acts as the third body. The recombination is essentially

a one-body process. A wall may, then, act as a catalyst for recombination, and surface catalysis is of great importance in nonequilibrium flows. We can write a surface catalysis rate equation as follows:

$$\left(\rho D \frac{\partial c}{\partial n}\right)_{\text{wall}} = \Gamma \sqrt{\frac{RT}{\pi m_2}} (c - c_{\text{eq}})_{\text{wall}}, \tag{46}$$

where the left-hand side is the rate of arrival of atoms at the surface by diffusion, Γ is the catalytic efficiency of the surface, the radical is a molecular velocity term, and the last bracket is the departure from equilibrium concentration at the surface. The rate of arrival is then seen to depend upon the degree of nonequilibrium at the surface. If $c = c_{\text{eq}}$ there is no net arrival of atoms, i.e., $\partial c/\partial n = 0$. The quantity Γ is usually not well known since the physical chemistry of catalytic reactions is not fully understood.

VII. **Viscous flows.** We have seen how chemical relaxation leads to the dispersion of waves, and other effects analogous to those due to viscosity. Often, relaxation and viscous processes must be considered together.

a. *Couette Flow.* Among problems of viscous, heating conducting flows with dissociation, the case of Couette flow is the simplest. Consider two infinite parallel plates separated by a distance δ, the lower plate being fixed and the upper plate moving at some velocity u_δ (which might be very high). In such a flow the shear force is constant, and the velocity and temperature profiles might be sketched in Figure 32. Many flows of interest can be represented, at least qualitatively, by suitably defined Couette flow situations. A variation of this problem is to hold both plates fixed and then impose a

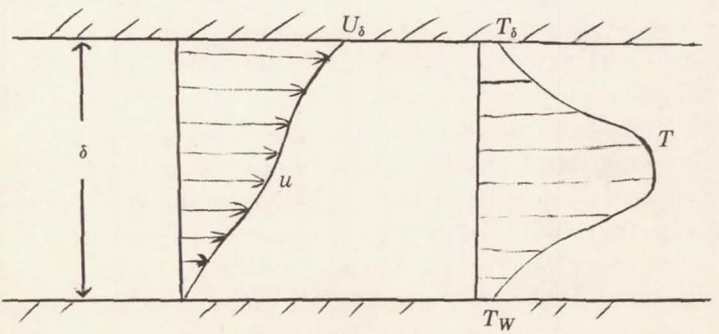


FIGURE 32. Couette Flow

temperature difference between the plates.

The differential equations are:

(47a) Momentum: $$\mu \frac{\partial u}{dy} = \text{const}$$

(47b) Energy: $$\mu\left[\Pr u \frac{du}{dy} + \frac{dH}{dy} + (\text{Le} - 1)\Pr h^{(0)} \frac{dc}{dy}\right] = \text{const}$$

(47c) Atom Production: $$\frac{d}{dy}\left(\text{Le}\,\mu \frac{dc}{dy}\right) = W_1.$$

Boundary conditions would specify velocity and temperature at the two surfaces as well as the catalytic condition

(47d) $$\left(\text{Le}\,\mu \frac{\mathrm{dc}}{dy}\right)_W = \Pr \Gamma \sqrt{\frac{RT_W}{\pi m_2}}(c - c_{\text{eq}})_W$$

at the wall.

Now, as in [**35**], one can eliminate μ between Equations (47a) and (47b) and integrate the resulting equation. Thus, if Le ≈ 1 and both walls are cold (Figure 31), one gets

(48a) $$\frac{H}{H_W} \approx 1 + \frac{1}{5}\Pr M_\delta^2\left(\frac{u}{u_\delta} - \frac{u^2}{u_\delta^2}\right).$$

Maximum enthalpy occurs when the expression in parentheses is 1/4, i.e., at $u/u_\delta = 1/2$. Then,

(48b) $$\left(\frac{H}{H_W}\right)_{\max} = 1 + \frac{\Pr M_\delta^2}{20}$$

and if

$$M_\delta = 25, \quad \left(\frac{H}{H_W}\right)_{\max} \approx 30 \text{ at } y = \frac{1}{2}\delta.$$

That is, the enthalpy at the center line is 30 times that at the plates. The foregoing indicates the large amount of heat which is introduced by viscous dissipation, most of which is absorbed in dissociation.

Consider now that the plates are both stationary, so that there is no dissipation. However, the upper one is at a very high temperature (see Figure 33). Equations (47) give relatively simple solutions for this problem, which is quite a good model for the conditions found at the stagnation region in hypersonic flow. Now, for Le = 1, $\mu(\partial H/\partial y)$ = const, and for the equilibrium case we may express a measure of

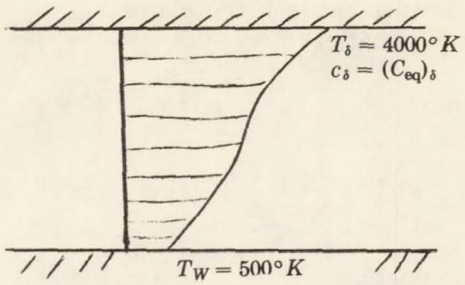


FIGURE 33. Plates at Different Temperatures

heat transfer as $Q \approx .65$ (this unit will remain an undefined "Nusselt number," serving only as a comparison for various cases of heat conduction). From Figure 33 we note that the temperature gradient is smallest midway between the plates because a greater proportion of the heat flux is due to diffusion of atoms in that region, while at the extreme surface temperatures heat is transferred chiefly by conduction.

Let us now examine the "frozen" case, where in this context, frozen means that the atom concentration is not changed by chemical reaction. That is, $W = 0$, and for this frozen case, $\mathrm{Le}\,\mu(\partial c/\partial y) = \text{const}$. This relation requires that the flux of atoms be constant at every layer. We consider that equilibrium exists at the upper plate, but not necessarily at the lower plate. Any of the straight lines in Figure 34 satisfy the condition $\partial c/\partial y = \text{const}$. (Assume $\mathrm{Le}\,\mu = \text{const}$, for convenience.) Suppose the lower surface is cold, so that $(c_{eq})_w = 0$; if the wall is completely noncatalytic, i.e., $\Gamma = 0$, then $\partial c/\partial y$ must vanish, according to Equation (46). If, on the other hand, the wall is fully catalytic, i.e., $\Gamma \to \infty$, then $c - c_{eq} \to 0$ for finite $\partial c/\partial n$. In the

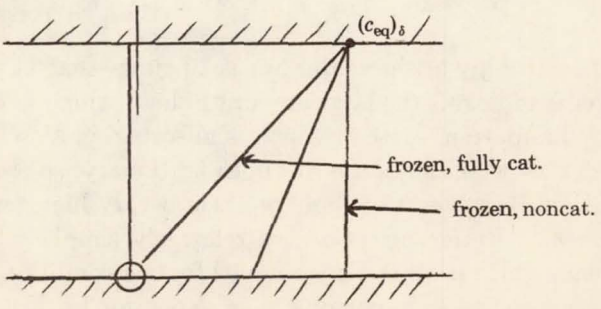


FIGURE 34. Temperature Distribution for the Frozen Case

frozen, noncatalytic case, the heat transfer (purely by conduction) is expected to be small, and using the unit introduced earlier, we find $Q \approx .3$. However, if the chemistry is frozen and the wall fully catalytic, there is a chemical reaction at the surface, that is, the energy of dissociation is deposited at the surface upon recombination, but there is no chemical activity in the gas. In this case, heat transfer is high and we get roughly $Q \approx 0.60$. Thus, a fully catalytic wall will experience about the same heat transfer whether or not there is gas-phase chemistry.

This kind of simple "Couette" calculation provides a sort of qualitative model for the more complete calculations of the hypersonic stagnation-point flow (see [28] and [31]).

b. *Couette Flow with Radiation.* Here, we consider that layers of the gas may exchange heat by radiation:

$$\frac{\mu}{\mathrm{Pr}} \frac{dH}{dy} + q_R = \text{const.}$$

Such a problem has been studied by Goulard and Goulard in [29] for Couette flow, with a very large temperature difference between the two plates. The gas is assumed to be quite transparent, $\delta \ll (1/\alpha_\nu)$ so that we do not have radiative equilibrium. However, we assume "quasi-equilibrium" and also assume a "gray gas." In this case the result from [29] is:

$$\frac{\mu}{\mathrm{Pr}} \frac{\partial H}{\partial y} + 2 \int_\tau^{\tau_\delta} B(\tau')d\tau' - 2 \int_0^\tau B(\tau')d\tau' = \text{const}$$

where $d\tau = \alpha dy$ and $B = \sigma T^4$; that is, B is B_ν integrated over all ν. The terms have the following meaning: $\partial H/\partial y$ represents heat conduction downward (Figure 35) from any point toward the low-temperature wall, the term $2\int_\tau^{\tau_\delta} B(\tau')d\tau'$ represents downward radiative transfer of heat for layers above τ, and the last term is upward radiation for the layers below τ. The net effect of radiative transfer is a smoothing out of the temperature profile. Temperature profiles and energy flux are shown in Figure 35 (taken from [29]). The presence of radiation revises the balance of heat flux. If the overall heat flux level is, say, a value of 3 watts/cm^2/sec., then at the upper plate, the analysis in [29] shows that 5 w/cm^2/sec. are due to downward convection, diminished by 2 w/cm^2/sec. due to upward radiation. At $y = 1/2$ there is a heat flux of 3 w/cm^2/sec. downward

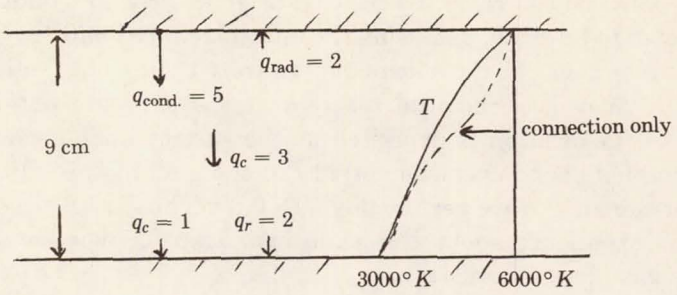


FIGURE 35. Couette Flow with Radiation

due to convection and the upward and downward radiation contribution tend to cancel each other. At the lower wall, there is then only $1\,w/cm^2/sec.$ downward due to convection and $2\,w/cm^2/sec.$ downward due to radiation. Chung in [**36**] has analyzed Couette flow with ionization, including effects of the plasma sheath.

c. *Viscous Waves.* A somewhat different kind of problem which provides a "model" for boundary layer flow is that of a single plate in a semi-infinite expanse of gas, the plate being suddenly moved or heated (see [**30**]). Here, we consider an unsteady problem in independent variables t and x to correspond to a two-dimensional steady problem. Since the plane surface is doubly infinite, all $\partial/\partial x = 0$, and the problem is linear: We may put all $D/Dt = \partial u/\partial t$.

If a step change of temperature or velocity occurs at the plate, a viscous, dissipative, conductive, or diffusive wave spreads into the gas above. In the absence of chemistry, the heat flux to the plate goes inversely with the wave thickness: $Q \propto 1/\sqrt{t}$. Thus, we will speak of heat transfer in terms of a Nusselt number $Q\sqrt{t}$ which would ordinarily be constant.

First suppose the temperature of the plate drops slightly, the gas above being dissociated. At first, there hasn't been time enough for any change in c (whatever the value of Γ) and the Nu is small, compared with the value later on, when c also falls and the chemical energy is given up to the surface, or near the surface in the gas. Ultimately, equilibrium must be achieved, for any Γ, though the value of Γ affects the speed of equilibration.

Figure 36 (taken from [**30**]) shows how Nu is affected by catalycity, as well as by ϵ which compares change in temperature with change in (equilibrium) concentration

$$\frac{\epsilon}{1+\epsilon} \equiv \left(\frac{Cp\,\Delta T}{Cp\,\Delta T + h^{(0)}\,\Delta c}\right)_{eq}$$

at constant pressure. ϵ can be quite small for air; in effect, because $h^{(0)}$ is large. Especially at low density is this true. The mathematical assumption of small ϵ greatly simplifies the analysis of this problem, providing a model which might usefully be extended to other problems (see [30]).

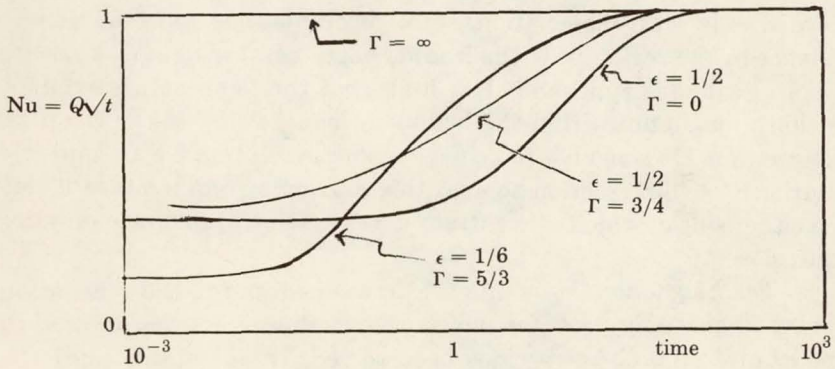


FIGURE 36. Effect of Catalycity

Figure 36 is remarkably close qualitatively, to the more exact results for stagnation-point flow due to Fay and Riddell in [31], and thus, this "Rayleigh Problem" provides a useful model for certain hypersonic boundary layers. Such a model for the flat plate boundary layer is provided by assuming that the plate temperature is low, but it is suddenly moved at high speed. Consequent dissipation produces heat flux. In this case, the early and late situations are the same—no chemistry. At intermediate times, however, atoms are produced in the gas, but may not have had time to recombine at the surface. Thus, at the surface, one has the concentration-history shown in Figure 37. A corresponding dip in an otherwise constant Nu may be expected at these intermediate times. Here again, small ϵ simplifies the analysis.

d. *The Leading Edge Problem.* The corresponding problem of the hypersonic boundary layer with chemistry at the leading edge of a flat plate has been studied by Rae in [33]. In this case, one is con-

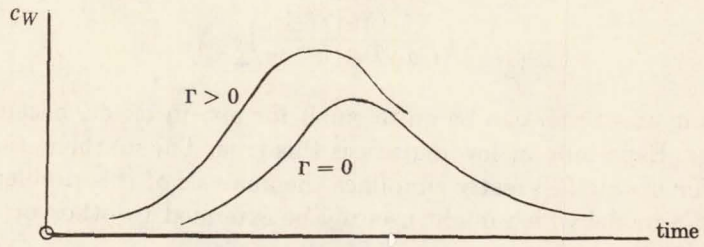


FIGURE 37. Concentration-History

cerned only with dissociation, since recombination is of no consequence in the region near the leading edge, i.e., $t \ll \tau_{rec}$. As a result, binary scaling is applicable. Rae finds that the temperature exhibits a sharp maximum within the boundary layer, owing to the effect of dissipation. He was able to achieve a simple solution by an approximation of a thin reaction zone at this maximum temperature layer, on either side of which concentration was assumed to change only by diffusion.

e. *The Stagnation Point.* As we have mentioned, the stagnation point is the only case for which the nonlinear viscous flow with nonequilibrium chemistry has been solved. In effect, one finds the leading term of a Taylor series in distance away from the stagnation point. The heat transfer results obtained in [31] are illustrated quite well by Figure 36, except that, instead of real time t, one uses the characteristic time L/U_∞ in forming the abscissa. Thus, a small nose radius would favor "frozen heat blockage," for a noncatalytic surface, for example.

f. *Stagnation Point Heat Transfer at Altitude.* Chung in [32] has analyzed this problem. For stagnation flows as altitude increases, τ increases, and this has the same effect as a decrease in L. Thus the flow tends to freeze, even though the boundary layer thickens. At high altitudes, the shock layer, δ is nearly all viscous (refer to Figure 17); that is, the boundary layer thickness approaches δ. The heat transfer rate calculated by Chung varies with altitude as shown in Figure 38 (taken from [32]) noncatalytic wall.

In region A, heat transfer decreases with increasing altitude due to slow recombination in the frozen boundary layer. The energy is essentially trapped in the form of dissociation energy, and the boundary layer, being almost completely frozen, has reduced heat transfer to the wall. However, as altitude increases still further, the

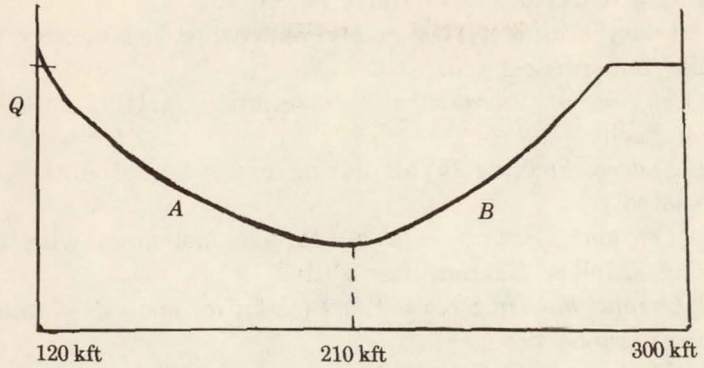


FIGURE 38. Heat Transfer Rate

degree of dissociation behind the shock decreases, and less and less energy is taken up in dissociation. Thus, at about 210 kilo feet these two effects are nearly equal and heat transfer reaches a minimum. Further increase in altitude, i.e., region B on Figure 38 results in failure to reach dissociation equilibrium behind the shock, and finally, at about 300 kilo feet, there is little or no energy going into dissociation at all. Here we approach the condition of an ideal gas, with no chemistry involved in the heat transfer process.

VIII. **Review of models discussed.** In these notes, a wide variety of assumptions and physical and mathematical models, useful in the analysis of hypersonic flows of a real dissociating gas, are discussed. In conclusion, these may be listed as follows:

1. *Lumped Constituents* (§Id): grouping together of nonreacting constituents.

2. *Lighthill Ideal Gas* (§If): vibration 50% excited and ρ_D constant; useful as a standard real gas.

3. *Nearly-Equal Speeds of Sound*: small dispersion of sound waves in either chemical (§IIa) or radiative (§IId) cases. Leads to *Telegraph Equation* (see §IIb).

4. *Quasi-Equilibrium Radiation* (§IIe): even if quite transparent, absorptivity and emissivity at black-body values.

5. *Gray Gas* (end of §IIe): Emissivity independent of wave length.

6. *Constant pressure, enthalpy behind normal shock* (end of §IIIb): For very strong plane shock, and γ nearly 1.

7. *Newtonian Flow* (§IVa): Enthalpy and velocity nearly constant,

for strong (curved) shocks and γ nearly 1.

8. *Binary Scaling* (§IIIc and §IVb): when low-density flow is initially underdissociated.

9. *Coupling of Vibration and Dissociation* (§IIIg): behind very strong shocks.

10. *Sudden Freezing* (§Va): during expansion of initially fully-dissociated gas.

11. *Constant "Entropy"* (§Vc): for channel flows with sudden freezing—Mollier diagram (see §Vd).

12. *Channel Flow in Stream Tubes* (§Vf): for analysis of nonequilibrium afterbody flows.

13. *Viscosity Evaluated in Hottest Part of Flow* (end of §VIa): generally a good rule. Viscosity *weakly dependent on c*.

14. *Constant* Pr, Le, $\rho\mu$ (§VIa): Usually done but not fully justified.

15. $Q \propto \Delta H$ (§VIa): Ideal gas result applies to this extent, if Le = 1.

16. *Couette Flow* (§VIIa): a model for hypersonic boundary layers.

17. *Viscous waves (Rayleigh Flow)* (§VIIc): a model for hypersonic boundary layers, especially stagnation point flow.

18. *Thin Reaction Layer* (§VIId): for leading-edge problem.

19. *Equilibrium Change of Internal Energy Much Smaller Than Chemical* (end of §VIIc): Small ϵ; Simplifies Rayleigh problem.

References

1. R. E. Meyerott, *Radiation heat transfer to hypersonic vehicles,* Lockheed Aircraft Corp. LMSD 2264-R1 (September 1958).

2. W. D. Hayes and R. F. Probstein, *Hypersonic flow theory,* Academic Press, New York, 1959.

3. D. L. Matthews, *Interferometric measurements in a shock tube of the dissociation rate of oxygen,* Phys. Fluids **2**(1959), 170.

4. S. R. Byron, *Interferometric measurement of the rate of dissociation of oxygen heated by strong shock waves,* Thesis, Cornell University, Ithaca, New York, 1958.

5. M. J. Lighthill, *Dynamics of a dissociating gas:* Part I, *Equilibrium flows,* J. Fluid Mech. **2**(1958), 1.

6. F. K. Moore, *Propagation of weak waves in a dissociated gas,* J. Aerospace Sci. **25**(1958), 279.

7. F. K. Moore and W. E. Gibson, *Propagation of weak disturbances in a gas subject to relaxation effects,* J. Aerospace Sci. **27**(1960), 117.

8. S. Chandrasekhar, *Radiative transfer,* Dover, New York, 1960.

9. M. J. Lighthill, *Dynamics of a dissociating gas:* Part II, *Quasi-equilibrium transfer theory,* J. Fluid Mech. **8**(1960), 161.

10. R. Goulard, *Fundamental equations of radiation gas dynamics,* AGARD Fluid Dynamics Panel Meeting, High-temperature Aspects of Hypersonic Flow (April 1962).

11. W. G. Vincenti and B. S. Baldwin, Jr., *Effect of thermal radiation on the propagation of plane acoustic waves,* J. Fluid Mech. **12**(1962), 449.

12. M. A. Heaslet and B. S. Baldwin, *Prediction of the structure of radiation-resisted shock waves,* Phys. Fluids **6**(1963), 781.

13. B. S. Baldwin, *The propagation of plane acoustic waves in a radiating gas,* NASA TR R-138(1962).

14. J. A. Morrison, *Wave propagation in Voight materials and visco-elastic materials with three parameter models,* Quart. Appl. Math. **14**(1956-1957), 153.

15. W. E. Gibson, *Dissociation scaling for nonequilibrium blunt nose flows,* ARS Journal **32**(1962), 285.

16. R. E. Duff and N. Davidson, *Calculation of reaction profiles behind steady state shock waves:* II, *The dissociation of air,* J. Chem. Phys. **31**(1959), 1018.

17. J. G. Hall, A. Q. Eschenroeder and P. V. Marrone, *Blunt-nosed inviscid air flows with coupled nonequilibrium processes,* J. Aerospace Sci. **29**(1962), 1038.

18. P. Hammerling, J. D. Teare and B. Kivel, *Theory of radiation from luminous shock waves in nitrogen,* Phys. Fluids **2**(1959), 422.

19. C. E. Treanor and P. V. Marrone, *Chemical relaxation with preferential dissociation from excited vibrational levels,* Cornell Aeronautical Laboratory Report QM-1626-A-10 (February 1963).

20. N. C. Freeman, *Nonequilibrium flow of an ideal dissociating gas,* J. Fluid Mech. **4**(1958), 407.

21. W. E. Gibson and P. V. Marrone, *Nonequilibrium blunt body flows: A correspondence between normal shock and blunt body flows,* Phys. Fluids **5**(1962), 1649; also *A similitude for nonequilibrium phenomena in hypersonic flight,* AGARD Meeting on High-temperature Aspects of Hypersonic Fluid Dynamics (March 1962). (To be published by Pergamon Press.)

22. W. Lick, *Invicid flow around a blunt body of a reacting mixture of gases,* Rensselaer Polytechnic Institute TR AE 5810 (May 1958) and TR AE 5814 (December 1958).

23. K. N. C. Bray, *Atomic recombination in a hypersonic wind tunnel nozzle,* J. Fluid Mech. **6**(1959), 1.

24. J. G. Hall and A. L. Russo, *Studies of chemical equilibrium in hypersonic nozzle flows,* Proc. First Conf. on Kinetics, Equilibria, and Performance of High Temperature Systems. Butterworth Sci. Pub. **219**(1960), 219-231.

25. K. N. C. Bray, *Simplified sudden freezing analysis for nonequilibrium nozzle flows,* ARS Journal **31**(1961), 831.

26. A. Q. Eschenroeder, *Entropy changes in nonequilibrium flows,* Cornell Aeronautical Laboratory Report AD-1689-A-2 (September 1962).

27. L. Lees, *Convective heat transfer with mass addition and chemical reactions,* Combustion and Propulsion Third AGARD Colloquium, Pergamon Press, 1958.

28. F. K. Moore (editor), *Theory of laminar flows.* Vol. IV, *High-speed aerodynamics and jet propulsion.* Princeton Univ. Press, Princeton, N. J., 1964.

29. R. Goulard and M. Goulard, *Energy transfer in the Couette flow of a radiant and chemically reacting gas,* Heat Trans. and Fluid Mech. Inst. Preprints of paper (Univ. of California) 1959.

30. F. K. Moore and W. J. Rae, *The Rayleigh problem for a dissociated gas,* Cornell Aeronautical Laboratory Report AF-1285-A-8 (June 1961); also *Hypersonic flow research,* Vol. 7 of Progress in Astronautics and Rocketry, edited by F. R. Riddell, Academic Press, New York, 1962, pp. 107-140.

31. J. A. Fay and F. R. Riddell, *Theory of stagnation point heat transfer in dissociated air,* J. Aerospace Sci. **25**(1958), 72.

32. P. M. Chung, *Hypersonic viscous shock layer of a nonequilibrium dissociating gas,* NASA TR T-109 (1961).

33. W. J. Rae, *An approximate solution for the nonequilibrium boundary layer near the leading edge of a flat plate,* Inst. Aero. Sci. Paper 62-178 (June 1962).

34. A. Q. Eschenroeder, D. W. Boyer and J. G. Hall, *Nonequilibrium expansions of air with coupled chemical reactions,* Phys. Fluids **5**(1962), 615.

35. J. F. Clarke, *Energy transfer through a dissociated diatomic gas in Couette flow,* J. Fluid Mech. **4**(1958), 441-465.

36. P. M. Chung, *Electrical characteristics of Couette and stagnation boundary layer flows of weakly ionized gases,* Aerospace Corp. Report No. TDR-169(3230-12) TN-2(1962).

37. W. L. Bade, E. A. Mason and K. S. Yun, *Transport properties of dissociated air,* ARS Journal **31** No. 4 (August 1961), 1151-1153.

38. F. K. Moore, *On the viscosity of dissociated air,* ARS Journal **32** No. 9 (September 1962), 1415.

CORNELL AERONAUTICAL LABORATORY
ITHACA, NEW YORK

P. J. Message

Decay of Orbits

I. **Expression for the drag force.**

a. *Idealized Case.* A body of mass m moves with speed V through a medium, of density ρ, at rest. In an interval of length δt, it captures the particles originally within a cylinder of length $V\delta t$, and cross section essentially equal to the cross section of the body (A, say). Thus the accreted mass is $AV\delta t\rho$, and if the increase in velocity is δV, we must have

$$mV = (m + AV\delta t\rho)(V + \delta V),$$

whence

$$m\frac{dV}{dt} = -AV^2\rho. \tag{1}$$

b. *Practical Assumption.* This derivation takes no account of the thermal motion of the medium. However the force (P) is found to be very nearly proportional to V^2 except for very small velocities, and we can put with close accuracy

$$P = -\frac{1}{2}C_D AV^2\rho, \tag{2}$$

where C_D, which is called the "drag coefficient." is close to 2 if the

mean free path in the medium is large compared to the dimensions of the body, becoming nearer to unity if the reverse is the case.

If the satellite is rotating steadily, the components of the force which are perpendicular to the relative velocity will cancel out almost entirely. If the rotation is about the axes of greatest moment of inertia, the mean value of A will remain constant. Ionization of the atmosphere is not likely to be important, since it is less than 5% at heights of less than 400 kilometers. In any case a change of C_D with time will not affect many of the results.

c. *Effect of Rotation of the Atmosphere.* This effect is small, so a simple approximation is justified. Let us suppose that the atmosphere rotates with the same angular velocity as the earth, Ω, say. Then its velocity at the point $\mathbf{r}$, where the origin is at the earth's center, is $\Omega \times \mathbf{r}$. If the satellite is at $\mathbf{r}$ with velocity $\mathbf{v}$, its velocity relative to the atmosphere is $\mathbf{V}$, where

$$\mathbf{V} = \mathbf{v} - \Omega \times \mathbf{r}. \tag{3}$$

Thus (writing $V = |\mathbf{V}|$, $v = |\mathbf{v}|$, etc.)

$$\begin{aligned} V^2 &= v^2 - 2\mathbf{v} \cdot (\Omega \times \mathbf{r}) + (\Omega \times \mathbf{r})^2 \\ &= v^2 - 2\Omega \cdot h + (\Omega \times \mathbf{r})^2 \\ &= v^2 - 2\Omega h \cos i + \Omega^2 r^2 \cos^2 \delta, \end{aligned}$$

where $\mathbf{h} = \mathbf{r} \times \mathbf{v}$ is the angular momentum per unit mass of the satellite about the earth's center, δ is its declination, and i the inclination of its orbit to the equator. We put

$$V^2 = v^2 F, \tag{4}$$

so that

$$F = 1 - \frac{2\Omega h \cos i}{v^2} + \frac{\Omega^2 r^2 \cos^2 \delta}{v^2}. \tag{4a}$$

The third term in F may usually be neglected, being less than about 1/250, and the second term, which is of the order of 1/15, may be evaluated at perigee, where the density is much greater than elsewhere on the orbit. Thus we may effectively regard F as constant.

Then the drag force is

$$\mathbf{P} = -\frac{1}{2} C_D A \rho V \mathbf{V} = -\frac{1}{2} C_D A \rho F v^2 \left(\frac{\mathbf{V}}{V}\right). \tag{5}$$

The component of $\mathbf{V}$ in the direction of the radius vector is

$$\frac{\mathbf{V}\cdot\mathbf{r}}{r}=\frac{\mathbf{v}\cdot\mathbf{r}}{r}=v\sin X, \quad \text{(using (3))}$$

where $\pi/2-X$ is the angle between the velocity and the radius vector. The component of $\mathbf{V}$ in the transverse direction in the orbit plane is

$$\begin{aligned}\frac{\mathbf{V}\cdot(\mathbf{h}\times\mathbf{r})}{hr}&=\{v^2r^2-(\mathbf{v}\cdot\mathbf{r})^2-(\boldsymbol{\Omega}\times\mathbf{r})-\mathbf{v}r^2\}/(hr) \quad \text{(using (3))}\\ &=\{v^2r^2\cos^2X-\Omega hr^2\cos i\}/(hr)\\ &=v\cos X-\Omega r\cos i \quad (\text{since } h=rv\cos X).\end{aligned}$$

The component of $\mathbf{V}$ in the direction of $\mathbf{h}$ is

$$\begin{aligned}\frac{\mathbf{V}\cdot\mathbf{h}}{h}&=-\frac{\boldsymbol{\Omega}\cdot(\mathbf{r}\times\mathbf{h})}{h} \qquad \text{(using (3))}\\ &=\{(\boldsymbol{\Omega}\cdot\mathbf{v})r^2-(\boldsymbol{\Omega}\cdot\mathbf{r})(\mathbf{r}\cdot\mathbf{v})\}/h.\end{aligned}$$

Now

$$\begin{aligned}\boldsymbol{\Omega}\cdot\mathbf{r}&=\Omega r\sin\delta\\ &=\Omega r\sin i\sin(\omega+f),\end{aligned}$$

where ω is the argument of perigee, and f the true anomaly. Also

$$\boldsymbol{\Omega}\cdot\mathbf{v}=\Omega v\sin i\cos(\omega+f-X),$$

so

$$\frac{\mathbf{V}\cdot\mathbf{h}}{h}=\Omega r\sin i\cos(\omega+f).$$

Therefore if $\hat{\mathbf{r}}$, $\hat{\mathbf{s}}$ and $\hat{\mathbf{h}}$ are unit vectors in the radial, transverse, and normal to orbit directions respectively, the drag force may be written

$$\mathbf{P}=-\frac{1}{2}C_DA\rho\sqrt{(F)}v^2\left\{\sin X\hat{\mathbf{r}}+\left(\cos X-\frac{\Omega r\cos i}{v}\right)\hat{\mathbf{s}}+\frac{\Omega r\sin i}{v}\cos(\omega+f)\hat{\mathbf{h}}\right\}.$$

Since $\mathbf{v}=v(\sin X\hat{\mathbf{r}}+\cos X\hat{\mathbf{s}})$, this may be written

$$\mathbf{P}=-\frac{m}{2}k\rho v\{\mathbf{v}-\Omega r\cos i\hat{\mathbf{s}}+\Omega r\sin i\cos(\omega+f)\hat{\mathbf{h}}\}, \tag{6}$$

where

$$k = \frac{C_D A \sqrt{F}}{m}, \tag{7}$$

m being the mass of the satellite.

II. The effect of the tangential component.

a. *The Equation of Motion.* The component of **P** parallel to the velocity **v** is the largest part. The equation of motion, considering only this part, is

$$\ddot{\mathbf{r}} = -\operatorname{grad} V - \frac{1}{2} k \rho v \mathbf{v}, \tag{7a}$$

where $V(\mathbf{r})$ is the gravitational potential.

b. *The Energy.* If $E = \frac{1}{2}v^2 + V$, then

$$\frac{dE}{dt} = \dot{\mathbf{v}} \cdot \mathbf{v} + \mathbf{v} \cdot \operatorname{grad} V = -\frac{1}{2} k \rho v^3. \tag{8}$$

Now $V = -\mu/r - R$, where R is the disturbing function for the oblateness, and so

$$E = \frac{1}{2}v^2 - \frac{\mu}{r} - R$$

$$= -\frac{\mu}{2a} - R.$$

$$\therefore \frac{\mu}{2a^2}\dot{a} - \dot{R} = -\frac{1}{2} k \rho v^3.$$

The change of the major semiaxis a due to the drag alone is therefore

$$\dot{a} = -\frac{k \rho a^2 v^3}{\mu}. \tag{9}$$

c. *The Angular Momentum.* If $\mathbf{h} = \mathbf{r} \times \mathbf{v}$,

$$\frac{d\mathbf{h}}{dt} = \mathbf{r} \times \ddot{\mathbf{r}} = \mathbf{r} \times \operatorname{grad} R - \frac{1}{2} k \rho v \mathbf{h}.$$

The change due to the drag alone is therefore

$$\frac{d\mathbf{h}}{dt} = -\frac{1}{2} k \rho v \mathbf{h}. \tag{10}$$

Therefore the direction of **h** is unaltered, and so the position of the orbit plane is unchanged by the tangential component.

Since $h = \{\mu a(1 - e^2)\}$, with (9),

$$\dot{e} = -\frac{1}{2e}k\rho(1 - e^2)v\left(\frac{av^2}{\mu} - 1\right)$$

$$= -\frac{1}{e}k\rho(1 - e^2)v\left(\frac{a}{r} - 1\right). \tag{11}$$

In terms of the eccentric anomaly E,

$$\dot{a} = -k\rho na^2\left(\frac{1 + e\cos E}{1 - e\cos E}\right)^{3/2}, \tag{9a}$$

and

$$\dot{e} = -k\rho na(1 - e^2)\cos E\frac{(1 + e\cos E)^{1/2}}{(1 - e\cos E)^{3/2}}. \tag{11a}$$

Hence the secular rates of change are

$$\bar{a} = -\frac{1}{2\pi}kna^2\int_0^{2\pi}\rho\frac{(1 + e\cos E)^{3/2}}{(1 - e\cos E)^{1/2}}dE, \tag{12}$$

and

$$\bar{\dot{e}} = -\frac{1}{2\pi}kna(1 - e^2)\int_0^{2\pi}\rho\cos E\left(\frac{1 + e\cos E}{1 - e\cos E}\right)^{1/2}dE. \tag{13}$$

King-Hele uses the quantity $x = ae$, whose secular rate of change is therefore

$$\bar{\dot{x}} = -\frac{1}{2\pi}kna^2\int_0^{2\pi}\rho(e + \cos E)\left(\frac{1 + e\cos E}{1 - e\cos E}\right)^{1/2}dE. \tag{14}$$

d. *King-Hele's First Order Theory of the Orbit.* In polar coordinates (r, θ) in the fixed plane of the motion the equations are

$$\ddot{r} - r\dot{\theta}^2 = -\frac{\mu}{r^2} - \frac{1}{2}k\rho v^2\sin X,$$

$$\frac{1}{r}\frac{d}{dt}(r^2\dot{\theta}) = -\frac{1}{2}k\rho v^2\cos X. \tag{15}$$

Putting $u = 1/r$, and $h = r^2\dot{\theta}$, and eliminating the time as independent variable, we obtain

$$\frac{d^2u}{d\theta^2} + u = \frac{\mu}{h^2} \tag{16}$$

and

$$\frac{dh}{d\theta} = -\frac{k\rho v}{2u^3}. \tag{17}$$

We use the Poisson method of succesive approximation, putting

$$\begin{aligned} u &= u_0 + \delta_1 u + \delta_2 u + \cdots; \\ h &= h_0 + \delta_1 h + \cdots. \end{aligned} \tag{18}$$

e. *Unperturbed Solution.*

$$\begin{aligned} &\frac{d^2u_0}{d\theta^2} + u_0 = \frac{\mu}{h_0^2} = \frac{1}{P_0}, \quad \text{say} \\ &\frac{dh_0}{d\theta} = 0. \end{aligned} \tag{19}$$

Whence h_0, and therefore P_0, is constant, and we obtain the equation of the ellipse

$$u_0 = \frac{1}{P_0}\{1 + e_0\cos(\theta - \omega_0)\}, \tag{20}$$

where e_0 and ω_0 are disposable constants.

f. *First Order Solution.* The equations are

$$\frac{d^2\delta_1 u}{d\theta^2} + \delta_1 u = -\frac{2\mu}{h_0^3}\delta_1 h \tag{21}$$

and

$$\frac{d\delta_1 h}{d\theta} = -\frac{k\rho v_0}{2u_0^3}, \tag{22}$$

where v_0 is v computed for the unperturbed orbit (20). From these we obtain

$$\frac{d^3\delta_1 u}{d\theta^3} + \frac{d\delta_1 u}{d\theta} = \frac{\mu k\rho V_0}{h_0^3 u_0^3} = g(\theta - \omega_0), \tag{23}$$

say. A particular solution is

$$\frac{d}{d\theta}(\delta_1 u) = \int_{\theta_0}^{\theta} g(\theta' - \omega_0)\sin(\theta - \theta')\, d\theta',$$

whence

$$\delta_1 u(\theta) = \int_{\theta_0}^{\theta} d\theta'' \int_{\theta_0}^{\theta''} d\theta' g(\theta' - \omega_0)\sin(\theta'' - \theta')$$

$$= \int_{\theta_0}^{\theta} d\theta' \int_{\theta'}^{\theta} d\theta'' g(\theta' - \omega_0)\sin(\theta'' - \theta')$$

$$= \int_{\theta_0}^{\theta} d\theta' g(\theta' - \omega_0)\{1 - \cos(\theta - \theta')\}.$$

The equations (12) and (14) may also be derived from this by considering the change in u from perigee to perigee, which leads to the expression

$$\Delta q = -q^2 \int_0^{2\pi} df g(f)(1 - \cos f) \tag{25}$$

for the change in one period of the perigee distance q and also the change in u from apogee to apogee, which leads to

$$\Delta q' = -q'^2 \int_0^{2\pi} df g(f)(1 + \cos f) \tag{26}$$

for the change in the apogee distance q'. Then transforming from the true anomaly f to the eccentric anomaly E, and use of $a = \frac{1}{2}(q + q')$, and $x = \frac{1}{2}(q' - q)$, leads to the previously found forms of equations (12) and (14).

III. **The form of the atmosphere.** We assume the surfaces of constant density to be ellipsoids of revolution, with the earth's rotational axis as axis of symmetry. If we assume the ellipticity of each ellipsoid to be the same as that of the geoid (King-Hele gives the value 0.003353), then at a height of 300 kilometers our ellipsoids remain within 1 kilometer distance of the ellipsoids in which the ellipticity varies to maintain hydrostatic equilibrium. Thus the polar equation of each surface may be written

$$r = r_0\{1 - \epsilon \sin^2\delta + O(\epsilon^2)\}, \tag{27}$$

where δ is the declination. Along a given radius vector, the density is nearly proportional to $\exp(-r/H)$, where H is the "scale height." If the perigee distance changes by more than about 100 kilometers, we must take account of the variation of H with altitude, but in fact this distance usually changes by less than 60 kilometers in the first 95% of the satellite's lifetime.

We will not consider the variations of density with time. Such changes do occur, with the period of the sun's axial rotation, with that of the earth's rotation (due to differences of day and night), and the sunspot cycle. These changes do not however affect the relative changes of a and x.

Thus we will take the density as

$$\rho = \rho_0 \exp\{-\beta r(1 + \epsilon \sin^2\delta)\}, \tag{28}$$

where $\beta = 1/H$.

IV. **King-Hele's treatment of the secular changes.**

a. *Approximate Equations of Motion.* Confining ourselves to small orbital eccentricities, we use the expansions

$$\frac{(1 + e\cos E)^{3/2}}{(1 - e\cos E)^{1/2}} = 1 + 2e\cos E + \frac{3}{4}e^2(1 + \cos 2E) + O(e^3)$$

and

$$(\cos E + e)\left(\frac{1 + e\cos E}{1 - e\cos E}\right)^{1/2} = \cos E + \frac{3}{2}e + \frac{1}{2}e\cos 2E + \frac{11}{8}e^2\cos E + \frac{1}{8}e^2\cos 3E + O(e^3).$$

Also

$$\sin^2\delta = \frac{1}{2}\sin^2 i\{1 - \cos 2(\omega + f)\}$$

$$= \frac{1}{2}\sin^2 i\{1 - \cos 2(\omega + E) + e\cos(2\omega + E) - e\cos(2\omega + 3E) + O(e^2)\}.$$

Hence the equations (12) and (14) yield, using also $r = a(1 - e\cos E)$,

$$\bar{a} = -\frac{1}{2\pi}na^2k\rho_0\exp(-\beta a)\int_0^{2\pi} dE\exp(\beta ae\cos E) \cdot\Big[1 + 2e\cos E + \frac{3}{4}e^2(1 + \cos 2E) - c\Big\{1 + e\cos E - \cos 2(\omega + E) - \frac{3}{2}e\cos(2\omega + E) + \frac{1}{2}e\cos(2\omega + 3E)\Big\} + O(e^3, ce^2)\Big], \tag{29}$$

and

$$\bar{\dot{x}} = -\frac{1}{2\pi} na^2 k\rho_0 \exp(-\beta a) \int_0^{2\pi} dE \exp(\beta ae \cos E)$$

$$\cdot \Big[\cos E + \frac{3}{2}e + \frac{1}{2}e\cos 2E + \frac{11}{8}e^2 \cos E + \frac{1}{8}e^2 \cos 3E$$

(30)

$$-c\Big\{ \cos E - \frac{1}{2}\cos(2\omega + E) - \frac{1}{2}\cos(2\omega + 3E) + e$$

$$-\frac{1}{2}e\cos 2\omega - e\cos 2(\omega + E) + \frac{1}{2}e\cos(2\omega + 4E)\Big\}$$

$$+ O(e^3, ce^2)\Big],$$

where $c = \frac{1}{2}\beta a\epsilon \sin^2 i$, which is usually less than 0.2.

b. *Bessel Functions of Imaginary Argument.* The functions $I_n(Z) = i^{-n}J_n(iZ)$ may be shown to satisfy the generating relation

$$\sum_{n=-\infty}^{\infty} I_n(Z)t^n = \exp\Big\{\frac{1}{2}Z\Big(t + \frac{1}{t}\Big)\Big\}, \tag{31}$$

from which, putting $t = \exp(i\varphi)$, we see that I_n must be

$$I_n(Z) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \exp(-in\varphi)\exp(Z\cos\varphi)\, d\varphi$$

(32)

$$= \frac{1}{2\pi}\int_0^{2\pi} \cos n\varphi \cdot \exp(Z\cos\varphi)\, d\varphi,$$

after a little rearrangement.

Also we see that

$$\int_0^{2\pi} \exp(Z\cos\varphi)\sin n\varphi\, d\varphi = 0. \tag{33}$$

So, putting $\beta ae = Z = ax$, equations (29) and (30) yield

$$\bar{a} = -na^2 k\rho_0 \exp(-\beta a)\Big[I_0 + 2I_1 e - I_0 c + \frac{3}{4}(I_0 + I_2)e^2$$

(34)

$$+ \Big\{ I_2 + \Big(\frac{3}{2}I_1 - \frac{1}{2}I_3\Big)e\Big\}c\cos 2\omega + O(e^3, ce^2)\Big],$$

and

$$\bar{\dot{x}} = -na^2k\rho_0\exp(-\beta a)\Big[I_1 + \frac{1}{2}(3I_0 + I_2)e - I_1c + \frac{1}{8}(11I_1 + I_3)e^2 + (I_2 - I_0)ce + \frac{1}{2}\{(I_1 + I_3) + (I_0 - I_4)e\}\cos 2\omega + O(e^3, ce^2)\Big]. \tag{35}$$

Division gives

$$\frac{\overline{da}}{dx} = \frac{I_0}{I_1} + 2e - \frac{1}{2}\left(\frac{I_0}{I_1}\right)\left(3\frac{I_0}{I_1} + \frac{I_2}{I_1}\right)e + \left(\frac{I_0}{I_1}\right)\left(\frac{I_2}{I_0} - \frac{1}{2} - \frac{1}{2}\frac{I_3}{I_1}\right) \cdot c\cos 2\omega + O(e^2, ce). \tag{36}$$

c. *Phase* 1. We have the asymptotic formula

$$I_n(Z) \sim \frac{\exp Z}{\sqrt{(2\pi Z)}}\left\{1 - \frac{4n^2 - 1^2}{1!\,8Z} + \frac{(4n^2 - 1^2)(4n^2 - 3^2)}{2!\,(8Z)^2} - \cdots\right\} \quad \text{as } Z \to \infty, \tag{37}$$

which is found to be useful if $Z > 3$, that is if $ae/H > 3$. Now since H is about 50 kilometers within about 50 %, and a is about 7000 kilometers, this means that this formula is useful if the eccentricity is greater than about 0.02. The part of the motion for which this is true is called "Phase I" by King-Hele. Use of this formula in (36) gives

$$\frac{da}{dx} = \frac{I_0}{I_1} - \frac{1}{\beta a} - \frac{3}{2\beta^2 ax} + \frac{e^2}{\beta x} - \frac{2c}{\beta^2 x^2}\cos 2\omega + O(e^2, ce). \tag{38}$$

Differentiation of (31) with respect to Z, and equating coefficients gives

$$\frac{dI_n}{dZ} = \frac{1}{2}(I_{n-1} + I_{n+1}), \tag{39}$$

and differentiation with respect to t similarly gives

$$nI_n = \frac{1}{2}Z(I_{n-1} - I_{n+1}), \tag{40}$$

and hence, on eliminating I_{n+1},

$$\frac{dI_n}{dZ} + \frac{n}{Z} I_n = I_{n-1}. \tag{41}$$

With $n = 1$, this gives

$$\frac{I_0}{I_1} = \frac{1}{Z} + \frac{1}{I_1}\frac{dI_1}{dZ}. \tag{42}$$

This enables equation (38) to be integrated to give (omitting the term in $\cos 2\omega$ for the moment)

$$\begin{aligned} a - a_0 = {} & \frac{1}{\beta}\ln\left(\frac{x}{x_0}\right) + \frac{1}{\beta}\ln\left\{\frac{I_1(\beta x)}{I_1(\beta x_0)}\right\} - \frac{x - x_0}{\beta a} \\ & + \frac{3}{2\beta^2 a_0}\ln\left(\frac{x}{x_0}\right) + O\left(\frac{xe}{\beta a}, ae^5, \frac{ae^2}{\beta^3 x^3}\right), \end{aligned} \tag{43}$$

suffix zero indicating initial values. Using the expansion for I_1, this gives, noting that the perigee distance q is given by $q = a - x$,

$$\begin{aligned} q - q_0 = {} & -\frac{1}{2}H\left[\left(1 - \frac{5H}{2a_0}\right)\ln\left(\frac{e_0}{e}\right)\right. \\ & \left. + (e_0 - e)\left\{\frac{3H(1 + e_0)}{4a_0 e e_0} - 1 + \frac{e + e_0}{2}\right\}\right] \\ & + O\left(ae^5, \frac{H^3}{a^2 e}\right). \end{aligned} \tag{44}$$

Returning to the term in $\cos 2\omega$, we note first that if we substitute (43) into (35) we obtain, for the dominant term,

$$\dot{x} = -\frac{K}{x} + \cdots,$$

where K is a constant. Thus

$$x^2 = x_0^2 - 2Kt + \cdots.$$

We know that, from the oblateness effect, the most important change of ω with time is a linear increase or decrease, according to whether the inclination is greater or less than the critical value. In either case, we may write to a first approximation,

$$\omega = A + Bx^2,$$

where A and B are constants.

We may now include the $\cos 2\omega$ term in our result. Integration of

this term leads to the addition to the right-hand side of equation (44) of the terms

$$-\frac{2cH^2}{x_0}\left(\cos 2\omega_0 - \frac{x_0}{x}\cos 2\omega\right)$$

$$\text{(44a)} \quad -4c\pi H^2\sqrt{(|B|/\pi)}\left\{\cos A\int_{2x_0\sqrt{(|B|/\pi)}}^{2x\sqrt{(|B|/\pi)}}\cos\left(\frac{1}{2}\pi\phi^2\right)d\phi + \frac{B}{|B|}\sin A\int_{2x_0\sqrt{(|B|/\pi)}}^{2x\sqrt{(|B|/\pi)}}\sin\left(\frac{1}{2}\pi\phi^2\right)d\phi\right\}.$$

Since the orbital period is given by $T/T_0 = (a/a_0)^{3/2}$, a relation similar to (44) may easily be derived between T, e, T_0, e_0 and H.

This relation, and the complete expression (44), have been used by King-Hele with success to derive H from the known values of q, T and e of a satellite's orbit at two stages in its lifetime, and from such determinations he has studied the variation in H with altitude and discovered its large changes as the sunspot cycle progresses.

d. *Phase* 2. For $e < 0.02$, we use the relations (39) through (42) to put equations (36) into the form

$$\beta\frac{da}{dZ} = \frac{1}{\beta Z} + \frac{1}{I_1}\frac{dI_1}{dZ} + \frac{1}{2\beta a}\frac{d}{dZ}\left\{4Z\frac{I_0}{I_1} - 6\ln(ZI_1)\right\} - \frac{0.88}{Z}c\cos(A + BH^2Z^2) + O(e^2, ec),$$

and from this is derived, by integrating and putting $q = a - x$,

$$\text{(45)} \quad q - q_1 = -H\Bigg[\left(1 - \frac{3H}{a_1}\right)\ln\left\{\frac{Z_1I_1(Z_1)}{ZI_1(Z)}\right\} + \frac{2H}{a_1}\left\{\frac{Z_1I_0(Z_1)}{I_1(Z_1)} - \frac{ZI_0(Z)}{I_1(Z)}\right\} - Z_1 + Z + 0.44c\left\{\cos A\int_{|B|H^2Z_1^2}^{|B|H^2Z^2}\frac{\cos\phi}{\phi}d\phi - \sin A\frac{B}{|B|}\int_{|B|H^2Z_1^2}^{|B|H^2Z^2}\frac{\sin\phi}{\phi}d\phi\right\} + O(e^2, 0.06c)\Bigg],$$

where the suffix "1" indicates initial values in Phase 2.

For a treatment of orbits of high eccentricity, see King-Hele's paper III.

V. **The effect on the apse, node, and inclination including the non-tangential components.** We will now take account of all of the components of the drag force. If we write R, S, and W for its components per unit mass in the radial, transverse and normal to orbit plane directions respectively, then from (6) we have

$$R = -\frac{1}{2}k\rho v^2 \sin X,$$

$$S = -\frac{1}{2}k\rho(v^2 \cos X - v\Omega r \cos i), \tag{46}$$

and

$$W = -\frac{1}{2}k\rho v\Omega r \sin i \cos(\omega + f).$$

The equation for the apse longitude is (see e.g. Brouwer and Clemence, p. 306),

$$\frac{d\varpi}{dt} = -\frac{b\cos f}{na^2 e}R + \frac{r(2 + e\cos f)\sin f}{nabe}S + 2\sin^2\left(\frac{1}{2}i\right)\frac{d\Omega}{dt}$$

$$= -\frac{k\rho v \sin f}{e}\left\{1 - \frac{1}{2}\frac{r^2\Omega\cos i}{h}(2 + e\cos f)\right\} \tag{47}$$

$$+ 2\sin^2\left(\frac{1}{2}i\right)\frac{d\Omega}{dt},$$

after some reductions, in which we make use of the relations $rv\cos X = h = nab$, and $v\sin X = (na^2/b)e\sin f$. The first term is an odd function of f, and so its mean value over an orbit is zero. The equation for the node longitude is

$$\frac{d\Omega}{dt} = \frac{r\sin(\omega + f)}{nab\sin t}W = -\frac{1}{4}\frac{k\rho v}{nab}\Omega r \sin 2(\omega + f). \tag{48}$$

This is also an odd function of f, and we conclude that the contributions of the drag to the secular motions of both the apse and the node are zero.

The equation for the orbital inclination is

$$\frac{di}{dt} = \frac{r\cos(\omega + f)}{nab}W = -\frac{1}{2}\frac{k\rho v}{nab}\Omega r^2 \sin i \cos^2(\omega + f), \tag{49}$$

which clearly has a strictly negative secular part, as found by Vinti

(1959), the leading term in expansion in powers of e being

$$-\frac{1}{4}k\rho_0 a\Omega \sin i \exp(-\beta a) I_0. \tag{50}$$

References

1. G. E. Cook, D. G. King-Hele and D. M. C. Walker, *Spherically symmetrical atmosphere,* Proc. Roy. Soc. Ser. A **257** (1960, 224-249.

2. ——, *With oblate atmosphere,* Proc. Roy. Soc. Ser. A **264** (1961), 88-121.

3. D. G. King-Hele, *High eccentricity orbits,* Proc. Roy. Soc. Ser. A **267** (1962), 541-557.

4. D. G. King-Hele and Janice Rees, *Scale height in the upper atmosphere, derived from changes in satellite orbits,* Proc. Roy. Soc. Ser. A **270** (1962), 562-587.

5. J. P. Vinti, *Theory of the effect of drag on the orbital inclination of an earth satellite,* J. Res. Nat. Bur. Standards **62** (1959), 79-88.

UNIVERSITY OF LIVERPOOL

Gen-ichiro Hori

The Effect of Radiation Pressure on the Motion of an Artificial Satellite

I. **Equations of motion; Disturbing function.** With the use of the Delaunay variables, L_1, L_2, L_3, L_4, l_1, l_2, l_3, l_4, which are usually denoted by L, G, H, K, l, g, h, k, respectively, the equations of motion of an artificial satellite around the oblate earth disturbed by solar radiation pressure are

$$\frac{dL_j}{dt} = \frac{\partial F}{\partial l_j}, \quad \frac{dl_j}{dt} = -\frac{\partial F}{\partial L_j} \qquad (j = 1, 2, 3, 4) \tag{1}$$

with the Hamiltonian

$$\begin{aligned} F = {} & \frac{\mu^2}{2L_1^2} - \nu L_4 + F_1(L_1, L_2, L_3, l_1, l_2, -, -) \\ & + F_2(L_1, L_2, L_3, l_1, l_2, -, -) + \Delta F_2(L_1, L_2, L_3, l_1, l_2, l_3, l_4), \end{aligned} \tag{2}$$

the motion of the sun being Keplerian.

In Equation (2), F_1, F_2 are severally the second and the fourth harmonics of the earth's potential, and ΔF_2, the contribution of solar radiation pressure. We assume that ΔF_2 is of the same order of magnitude as F_2, and $-\nu L_4$ is assumed to be of the first order. The system (1) has four degrees of freedom but the Hamiltonian is free from the time. The fourth angular variable l_4 stands for mean longi-

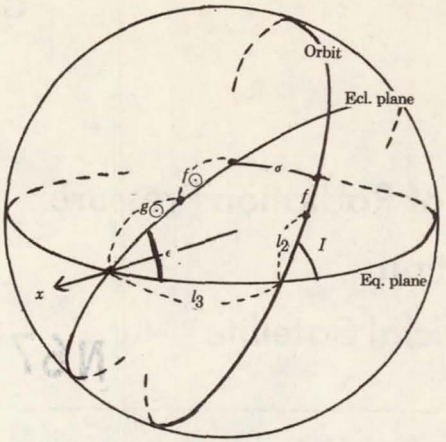


FIGURE 1. Satellite Orbit

tude of the sun, νt + const, so that $-\nu L_4$ in Equation (2) may represent the kinetic energy of the sun.

If σ denotes the angular distance between the sun and the satellite; r, f, the radius vector and the true anomaly of the satellite; $r_\odot$, $f_\odot$, the corresponding quantities of the sun, we have

$$\Delta F_2 = -\beta \frac{r}{r_\odot^2} \cos \sigma.$$

If further x, y, z; $x_\odot$, $y_\odot$, $z_\odot$ stand for their rectangular coordinates with the xy-plane in the equatorial plane and the x-axis pointing to the vernal equinox, we have

$$\cos \sigma = \frac{x}{r} \cdot \frac{x_\odot}{r_\odot} + \frac{y}{r} \cdot \frac{y_\odot}{r_\odot} + \frac{z}{r} \cdot \frac{z_\odot}{r_\odot}.$$

But (refer to Figure 1)

$$\frac{x}{r} = \cos (f + l_2) \cos l_3 - \sin (f + l_2) \sin l_3 \cos I,$$

$$\frac{y}{r} = \cos (f + l_2) \sin l_3 + \sin (f + l_2) \cos l_3 \cos I,$$

$$\frac{z}{r} = \sin (f + l_2) \sin I;$$

$$\frac{x_\odot}{r_\odot} = \cos(f_\odot + g_\odot),$$

$$\frac{y_\odot}{r_\odot} = \sin(f_\odot + g_\odot)\cos\epsilon,$$

$$\frac{z_\odot}{r_\odot} = \sin(f_\odot + g_\odot)\sin\epsilon,$$

then,

$$\Delta F_2 = -\beta \frac{a}{a_\odot^2} \cdot \frac{r}{a} \left(\frac{a_\odot}{r_\odot}\right)^2 (A\cos f + B\sin f),$$

$$A = \frac{1}{4}(1+\theta)(1+\theta_\odot)\cos(l_2 + l_3 - f_\odot - g_\odot)$$

$$+\frac{1}{4}(1+\theta)(1-\theta_\odot)\cos(l_2 + l_3 + f_\odot + g_\odot)$$

$$+\frac{1}{4}(1-\theta)(1+\theta_\odot)\cos(l_2 - l_3 + f_\odot + g_\odot)$$

$$+\frac{1}{4}(1-\theta)(1-\theta_\odot)\cos(l_2 - l_3 - f_\odot - g_\odot)$$

$$+\frac{1}{2}\sin I \sin\epsilon\,[\cos(l_2 - g_\odot - f_\odot) - \cos(l_2 + f_\odot + g_\odot)],$$

$$B = -\frac{1}{4}(1+\theta)(1+\theta_\odot)\sin(l_2 + l_3 - f_\odot - g_\odot)$$

$$-\frac{1}{4}(1+\theta)(1-\theta_\odot)\sin(l_2 + l_3 + f_\odot + g_\odot)$$

$$-\frac{1}{4}(1-\theta)(1+\theta_\odot)\sin(l_2 - l_3 + f_\odot + g_\odot)$$

$$-\frac{1}{4}(1-\theta)(1-\theta_\odot)\sin(l_2 - l_3 - f_\odot - g_\odot)$$

$$-\frac{1}{2}\sin I \sin\epsilon\,[\sin(l_2 - f_\odot - g_\odot) - \sin(l_2 + f_\odot + g_\odot)],$$

where a stands for the semi-major axis and

$$\theta = \cos I, \qquad \theta_\odot = \cos\epsilon,$$

the subscript $\odot$ referring to the sun.

We may assume with enough accuracy that the sun moves in a circular orbit. Under this assumption $f_\odot + g_\odot$ reduces to l_4, and we have

$$\Delta F_2 = -\beta \frac{a}{a_\odot^2} \cdot \frac{r}{a} (A \cos f + B \sin f), \tag{3}$$

$$\begin{aligned} A = {} & \frac{1}{4}(1+\theta)(1+\theta_\odot)\cos(l_2+l_3-l_4) \\ & + \frac{1}{4}(1+\theta)(1-\theta_\odot)\cos(l_2+l_3+l_4) \\ & + \frac{1}{4}(1-\theta)(1+\theta_\odot)\cos(l_2-l_3+l_4) \\ & + \frac{1}{4}(1-\theta)(1-\theta_\odot)\cos(l_2-l_3-l_4) \\ & + \frac{1}{2}\sin I \sin\epsilon[\cos(l_2-l_4) - \cos(l_2+l_4)], \end{aligned} \tag{4}$$

and a similar expression for B which we do not need in the present discussion.

II. **Elimination of short period terms.** Since ΔF_2 is assumed of the second order, this has no contribution to short period terms of the first order. We may have, after eliminating l_1 from the Hamiltonian, the new equations of motion,

$$\frac{dL_j'}{dt} = \frac{\partial F^*}{\partial l_j'}, \quad \frac{dl_j'}{dt} = -\frac{\partial F^*}{\partial L_j'} \qquad (j = 1, 2, 3, 4) \tag{5}$$

with

$$\begin{aligned} F^* = {} & F_0^*(L_1') - \nu L_4' + F_1^*(L_1', L_2', L_3') \\ & + F_2^*(L_1', L_2', L_3', -, l_2', -, -) \\ & + \Delta F_2^*(L_1', L_2', L_3', -, l_2', l_3', l_4'), \end{aligned} \tag{6}$$

where F_0^* stands for $\mu^2/2L_1'^2$ and L_3', l_4' are identical with L_3, l_4 respectively.

In Equation (6), F_1^*, F_2^* are provided by any theory of the main problem of artificial satellite motion. (See the article by D. Brouwer, Astronom. J. **64** (1959), 378-397). It is a principal feature of the main problem that F_1^* is free from l_2' or g' and then free from all the angular variables. F_1^* has been found to be

$$F_1^* = \frac{\mu^4 k_2}{L_1'^3 L_2'^3}\left(-\frac{1}{2}+\frac{3}{2}\frac{L_3'^2}{L_2'^2}\right). \tag{7}$$

On the other hand, the relations

$$\left[\frac{r}{a}\cos f\right]_s = -\frac{3}{2}e, \qquad \left[\frac{r}{a}\sin f\right]_s = 0,$$

the subscript s denoting the constant part in l_1 or l, give

$$\Delta F_2^* = \frac{3}{2}\beta\frac{ae}{a_\odot^2}A, \tag{8}$$

$$\begin{aligned} A = {} & \frac{1}{4}(1+\theta)(1+\theta_\odot)\cos(l_2'+l_3'-l_4') \\ & +\frac{1}{4}(1+\theta)(1-\theta_\odot)\cos(l_2'+l_3'+l_4') \\ & +\frac{1}{4}(1-\theta)(1+\theta_\odot)\cos(l_2'-l_3'+l_4') \\ & +\frac{1}{4}(1-\theta)(1-\theta_\odot)\cos(l_2'-l_3'-l_4') \\ & +\frac{1}{2}\sin I\sin\epsilon\left[\cos(l_2'-l_4')-\cos(l_2'+l_4')\right], \end{aligned} \tag{9}$$

where

$$\begin{aligned} \theta = \frac{L_3'}{L_2'}, \quad \sin I = \sqrt{\left(1-\frac{L_3'^2}{L_2'^2}\right)}, \quad \theta_\odot = \cos\epsilon, \\ a = \frac{L_1'^2}{\mu}, \qquad e = \sqrt{\left(1-\frac{L_2'^2}{L_1'^2}\right)}. \end{aligned} \tag{10}$$

III. **Elimination of long period terms; Nonresonance case.** Consider a canonical transformation $L_j', l_j' \to L_j'', l_j''$ with a determining function

$$S^* = \sum_{j=1}^{4} L_j'' l_j' + S_1^*(L_1'', L_2'', L_3'', -, l_2', l_3', l_4'):$$

(11)
$$L_j' = \frac{\partial S^*}{\partial l_j'} = L_j'' + \frac{\partial S_1^*}{\partial l_j'}, \quad l_j'' = \frac{\partial S^*}{\partial L_j''} = l_j' + \frac{\partial S_1^*}{\partial L_j''}.$$

The requirement that a new Hamiltonian F^{**} should be free from all the angular variables l_j' $(j = 1, 2, 3, 4)$ may yield

$$F_0^*(L_1'') - \nu \times \left(L_4'' + \frac{\partial S_1^*}{\partial l_4'}\right) + F_1^*\left(L_1'', L_2'' + \frac{\partial S_1^*}{\partial l_2'}, L_3'' + \frac{\partial S_1^*}{\partial l_3'}\right)$$
$$+ F_2^*(L_1'', L_2'', L_3'', -, l_2', -, -) + \Delta F_2^*(L_1'', L_2'', L_3'', -, l_2', l_3', l_4')$$
$$= F_0^{**} + F_1^{**} + F_2^{**}.$$

order 0: $$F_0^{**} = F_0^*(L_1'') = \frac{\mu^2}{2L_1''^2};$$

order 1: $$F_1^{**} = -\nu L_4'' + F_1^*(L_1'', L_2'', L_3'');$$

order 2: $$-\nu \frac{\partial S_1^*}{\partial l_4'} + \frac{\partial F_1^*}{\partial L_2''}\frac{\partial S_1^*}{\partial l_2'} + \frac{\partial F_1^*}{\partial L_3''}\frac{\partial S_1^*}{\partial l_3'} + F_2^* + \Delta F_2^* = F_2^{**},$$

or, if the subscripts s and p refer severally to constant and periodic parts in l_2', l_3', l_4',

(12) $$F_2^{**} = F_{2s}^*(L_1'', L_2'', L_3''),$$

(13) $$\nu \frac{\partial S_1^*}{\partial l_4'} - \frac{\partial F_1^*}{\partial L_2''}\frac{\partial S_1^*}{\partial l_2'} - \frac{\partial F_1^*}{\partial L_3''}\frac{\partial S_1^*}{\partial l_3'} = F_{2p}^* + \Delta F_2^*.$$

When none of the combinations $\nu + i(\partial F_1^*/\partial L_2'') + j(\partial F_1^*/\partial L_3'')$ $(i, j = 0, \pm 1)$, vanishes, or remains of the first order (nonresonance case), integration of Equation (13) results in

$$S_1^* = S_{1,\mathrm{obl}}^* + \Delta S_1^*,$$

where $S_{1,\mathrm{obl}}^*$ is the contribution of F_{2p}^* and given by the theory of the main problem; ΔS_1^* is the contribution of solar radiation pressure. Let n_2, n_3 be

$$
\begin{aligned}
n_2 &= -\frac{\partial F_1^*}{\partial L_2''} = nJ_2\left(\frac{R}{a}\right)^2 \eta^{-4}\left(-\frac{3}{4}+\frac{15}{4}\theta^2\right), \\
n_3 &= -\frac{\partial F_1^*}{\partial L_3''} = nJ_2\left(\frac{R}{a}\right)^2 \eta^{-4}\left(-\frac{3}{2}\theta\right),
\end{aligned} \tag{14}
$$

where R stands for the equatorial radius of the earth, J_2, the coefficient of the second harmonic of the earth's potential when developed in the Vinti form ($J_2 = +1.082 \cdot 10^{-3}$) and $n = \mu^2/L_1''^3$, the mean motion of mean anomaly l_1, and $\eta = L_2''/L_1''$. We then have

$$
\begin{aligned}
\Delta S_1^* = \frac{3}{2}\beta\frac{ae}{a_\odot^2}\Bigg[& \frac{1}{4}(1+\theta)(1+\theta_\odot)\frac{\sin(l_2'+l_3'-l_4')}{n_2+n_3-\nu} \\
& +\frac{1}{4}(1+\theta)(1-\theta_\odot)\frac{\sin(l_2'+l_3'+l_4')}{n_2+n_3+\nu} \\
& +\frac{1}{4}(1-\theta)(1+\theta_\odot)\frac{\sin(l_2'-l_3'+l_4')}{n_2-n_3+\nu} \\
& +\frac{1}{4}(1-\theta)(1-\theta_\odot)\frac{\sin(l_2'-l_3'-l_4')}{n_2-n_3-\nu} \\
& +\frac{1}{2}\sin I \sin\epsilon\left\{\frac{\sin(l_2'-l_4')}{n_2-\nu}-\frac{\sin(l_2'+l_4')}{n_2+\nu}\right\}\Bigg],
\end{aligned} \tag{15}
$$

a, e, θ being given by Equations (10) but with L_1', L_2', L_3' replaced by L_1'', L_2'', L_3'' respectively. Since β is assumed $O(J_2^2)$, and ν, $O(nJ_2)$, ΔS_1^* is $O(J_2)$ as expected. Equation (15) yields the long period terms of the first order due to solar radiation pressure in the nonresonance case. Solar radiation pressure has no contribution to secular terms in this case.

IV. **The condition for resonance.** When some of the denominators in Equation (15) vanish or reduce to the second order, Equation (15) looses its validity (resonance case). From the relation, $n^2a^3 = \mu$, we have

$$
\nu = n \cdot \nu\mu^{-1/2} R^{3/2}\left(\frac{a}{R}\right)^{3/2}, \tag{16}
$$

or

$$\nu = \nu_0 \left(\frac{a}{R}\right)^{3/2} n$$

with

$$\nu_0 = \nu\mu^{-1/2} R^{3/2} = 0.1606 \cdot 10^{-3},$$

ν_0 being the ratio of ν to the mean motion of a fictitious satellite with R as its semi-major axis.

Then Equations (14), (16) give

$$n_2 + in_3 + j\nu = nJ_2 \left(\frac{R}{a}\right)^2 \left[\eta^{-4}\left(-\frac{3}{4} - i\frac{3}{2}\theta + \frac{15}{4}\theta^2\right) + j\frac{\nu_0}{J_2}\left(\frac{a}{R}\right)^{7/2}\right] \tag{17}$$

$$(i = 0, \pm 1;\ j = \pm 1),$$

hence the condition for resonance is

$$\frac{15}{4}\theta^2 - i\frac{3}{2}\theta - \frac{3}{4} + j\frac{\nu_0}{J_2}\left(\frac{a}{R}\right)^{7/2}\eta^4 = 0, \tag{18}$$

or

$$\left(\frac{a}{R}\right)^{7/2}\eta^4 = j\frac{3}{4}\frac{J_2}{\nu_0}(1 + i2\theta - 5\theta^2). \tag{18'}$$

This is a relation between a, $e(=\sqrt{(1-\eta^2)})$ and θ.

If $(a/R)^{7/2}\eta^4$ is measured along an ordinate, and θ along an abscissa, Equation (18) or (18′) represents six parabolas of the same shape with their axes parallel to the ordinate. (See Figure 2.)

V. **Resonance cases.** When a set of values of a, e, θ is such that it lies on any of the six parabolic arcs in Figure 2, we have a resonance of simple character, and may treat it in a way similar to the case of the critical inclination in the main problem. Assume the mean motion of an argument $l_2' + il_3' + jl_4'$, that is, $n_2 + in_3 + j\nu$ vanishes or reduces to $O(J_2^2)$, and let $Q_2 \cos(l_2' + il_3' + jl_4')$ be the resonance term in ΔF_2^*. In place of Equation (12), (13), we may have

$$F_2^{**} = F_{2s}^*(L_1'', L_2'', L_3'') + Q_2(L_1'', L_2'', L_3'')\cos(l_2' + il_3' + jl_4'), \tag{19}$$

$$\nu\frac{\partial S_1^*}{\partial l_4'} - \frac{\partial F_1^*}{\partial L_2''}\frac{\partial S_1^*}{\partial l_2'} - \frac{\partial F_1^*}{\partial L_3''}\frac{\partial S_1^*}{\partial l_3'} = F_{2p}^* + \Delta F_2^* - Q_2\cos(l_2' + il_3' + jl_4'). \tag{20}$$

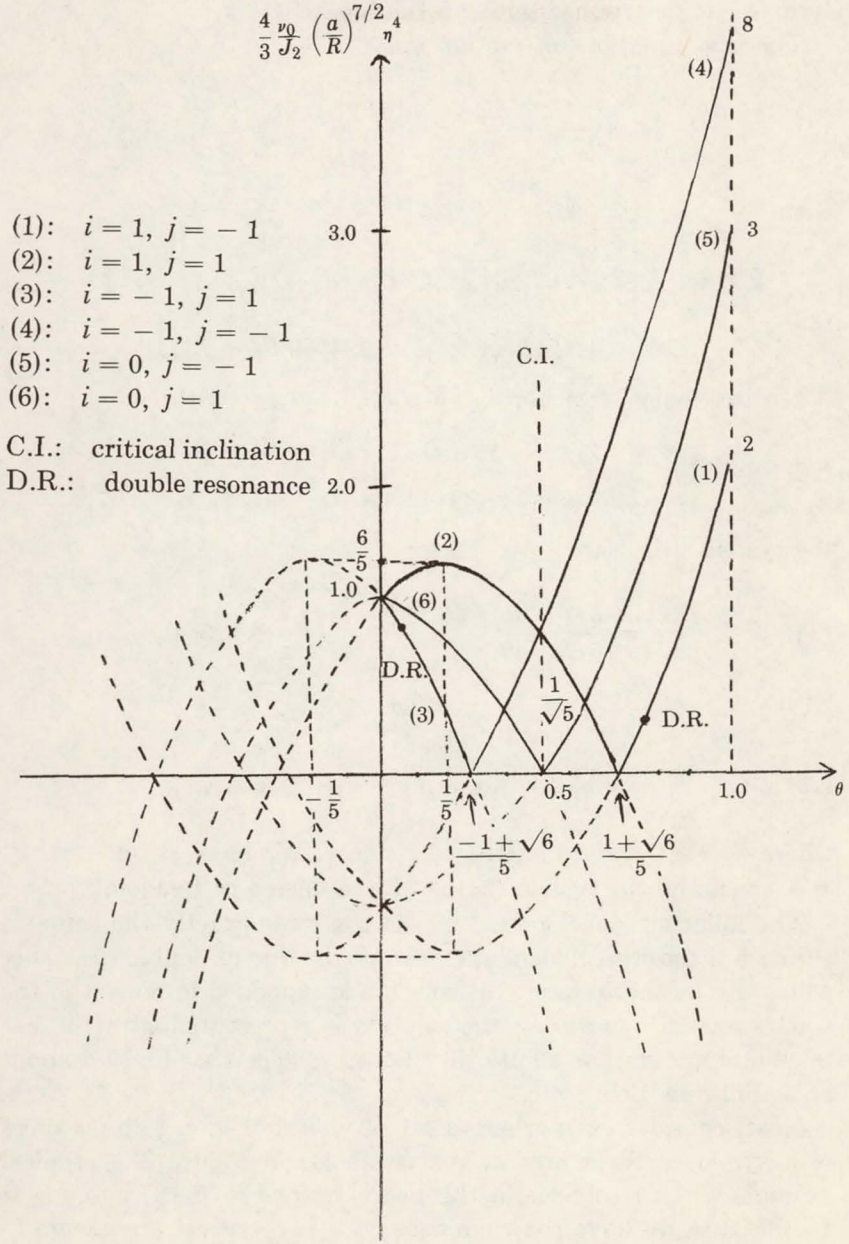


FIGURE 2. Conditions for Resonance

Since the right hand member of Equation (20) has no resonance term, S_1^* is determined with no trouble.

The new equations of motion are

$$\frac{dL_j''}{dt} = \frac{\partial F^{**}}{\partial l_j''}, \quad \frac{dl_j''}{dt} = -\frac{\partial F^{**}}{\partial L_j''} \qquad (j = 1, 2, 3, 4) \tag{21}$$

with

$$\begin{aligned} F^{**} = {} & \frac{\mu^2}{2L_1''^2} - \nu L_4'' + F_1^*(L_1'', L_2'', L_3'') + F_{2s}^*(L_1'', L_2'', L_3'') \\ & + Q_2(L_1'', L_2'', L_3'') \cos(l_2'' + il_3'' + jl_4''). \end{aligned} \tag{22}$$

After a canonical transformation $L_j'', l_j'' \to x_j, y_j$ with

$$\begin{aligned} & y_1 = l_1'', \ y_2 = l_2'', \ y_3 = l_3'', \ y_4 = l_2'' + il_3'' + jl_4''; \\ & x_1 = L_1'', \ x_2 = L_2'' - jL_4'', \ x_3 = L_3'' - ijL_4'', \ x_4 = jL_4'', \end{aligned} \tag{23}$$

the system (21) leads to

$$\frac{dx_j}{dt} = \frac{\partial \overline{F}}{\partial y_j}, \quad \frac{dy_j}{dt} = -\frac{\partial \overline{F}}{\partial x_j} \qquad (j = 1, 2, 3, 4), \tag{24}$$

with

$$\overline{F} = \frac{\mu^2}{2x_1^2} - j\nu x_4 + \overline{F}_1 + \overline{F}_2 + Q_2 \cos y_4, \tag{25}$$

where $\overline{F}_1$, $\overline{F}_2$, Q_2 are functions of x's only. We have at once, x_1, x_2, $x_3 = \text{const}$ and the system (24) is of one degree of freedom.

The following development of the theory is exactly the same as the case of the critical inclination except the case of double resonance which will be shown later: the solution is expanded in powers of the square root of J_2 or β/J_2; the solution is represented with the use of elliptic integrals or elliptic functions; y_4 makes a libration about an equilibrium point, etc.

Next, consider cases where a set of values of a, e, θ fits a cross of any two or three arcs of the parabolas in Figure 2. A typical example is the resonance in the polar orbit, $\theta = O(J_2)$, and $j = 1$. In this case we have the resonance with two critical arguments l_3'' and $l_2'' + l_4''$. An entirely different approach is required.

VI. **Double resonance.** Returning to the system (24), we may find that the development of the theory in powers of the square root of J_2 or β/J_2 becomes also false when $\partial^2 \bar{F}_1/\partial x_4^2$ vanishes (double resonance).

From

$$\bar{F}_1 = \frac{\mu^4 J_2 R^2}{x_1^3 (x_2 + x_4)^3}\left[-\frac{1}{4} + \frac{3}{4}\frac{(x_3 + ix_4)^2}{(x_2 + x_4)^2}\right],$$

we have

$$\frac{\partial^2 \bar{F}_1}{\partial x_4^2} = \frac{3}{2}\frac{\mu^4 J_2 R^2}{x_1^3 (x_2 + x_4)^5}(i^2 - 2 - i \cdot 10\theta + 15\theta^2). \tag{26}$$

Since $i = 0$ occurs only in polar orbits, the double resonance occurs if

$$15\theta^2 - 10\theta - 1 = 0 \text{ or } \theta = \frac{1}{3} + \frac{2}{15}\sqrt{10} = 0.754 \qquad (\text{for } i = 1)$$

or

$$15\theta^2 + 10\theta - 1 = 0 \text{ or } \theta = -\frac{1}{3} + \frac{2}{15}\sqrt{10} = 0.088 \qquad (\text{for } i = -1).$$

In this case the solution may be developed in powers of the cube root of J_2 or β/J_2 which is very ineffective for practical application. We need a new approach.

VII. **A remark.** Insofar as we use a method of successive approximations, the most difficult case looks like the case where $n_2 + in_3 + j\nu$

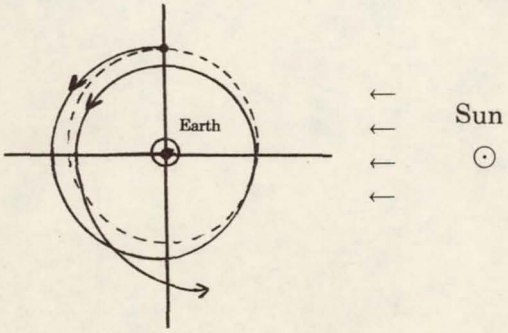


FIGURE 3. Decreasing Perigee Distance

is identically zero for any values of i and j: it is a case of ∞-ple resonance. This case is, however, nothing but the case where the earth is supposed to be a sphere and the sun is assumed to be fixed in space. The case, however, permits an exact solution if we use parabolic coordinates (when the sun is fixed at infinity) or elliptic coordinates (when the sun is fixed at a finite distance). In the two-dimensional problem the perigee distance decreases secularly until the satellite collides with the earth. (See Figure 3.)

UNIVERSITY OF TOKYO
TOKYO, JAPAN

S. V. Parter

Special Computation Procedures for Differential Equations

I. **Introduction.** Many problems of applied mathematics arise naturally as differential equations. In most cases there is no hope of finding an explicit, closed representation of the solution. Thus we are led to the computer. However, the availability of high-speed computers does *not* mean that "practical men" can give up the analytical study of differential equations.

Indeed, in some sense, the great advance in our computational ability requires that we put more effort into the analytical study. After all, twenty years ago we could only shrug our shoulders at these problems. Now we *can* and *do* attempt to get approximate results. And, in order to get computational results that are meaningful, we must do some analysis.

In these lectures, I hope to present some of the ideas and results in this area.

II. **Ordinary differential equations.** The simplest problem is the *Pure Initial-Value Problem*

$$y' = f(x,y), \qquad y(x_0) = y_0. \tag{II.1}$$

Here $y = y(x)$ may be a vector and then (II.1) represents a system of equations. It is well known that almost every Initial-Value problem may be put in this form. For example, suppose we start with

$$\text{(II.2)} \qquad \begin{aligned} y^{(\nu)} &= f(x, y, y', \cdots, y^{(\nu-1)}), \\ y^{(j)}(x_0) &= y_0^j, \qquad j = 0, 1, \cdots, \nu - 1. \end{aligned}$$

Then we set $Z_1 = y$, $Z_2 = y'$, $\cdots$, $Z_\nu = y^{(\nu-1)}$, and we write (II.2) as

$$\text{(II.1}') \qquad \begin{aligned} Z_1' &= Z_2, \\ Z_2' &= Z_3, \\ &\cdots \\ Z_{\nu-1}' &= Z_\nu, \\ Z_\nu' &= f(x, Z_1, Z_2, \cdots, Z_\nu), \\ Z_j(x_0) &= y_0^{j-1}, \qquad j = 1, 2, \cdots, \nu. \end{aligned}$$

We now turn to the question of numerical methods for approximating the solution $y(x)$ of (II.1).

We assume that $f(x, y)$ is continuous in (x, y) and satisfies a Lipschitz condition in y, i.e., there is a constant L such that

$$\text{(II.3)} \qquad \|f(x, y) - f(x, z)\| \leqq L\|y - z\|.$$

This condition assures us of the existence of a unique solution $y(x)$ in some neighborhood of x_0. For those who are skeptical of such mathematical niceties, let us consider two examples.

EXAMPLE 1. $y' = y^{1/2}$, $y(0) = 0$.

Then $y_1(x) \equiv 0$ and $y_2(x) = \frac{1}{4}x^2$ are solutions in the interval $0 \leqq x \leqq 1$. Here we do not have a unique solution.

EXAMPLE 2. $y' = 1 + y^2$, $y(0) = 0$.

In this case $y(x) = \tan x$ and there is no solution at all in the "larger" interval $0 \leqq x \leqq \pi$.

Now, let an increment h be chosen; then we seek values Y_k which approximate $y(kh + x_0)$.

The simplest formula we can use is

$$\text{(II.4)} \qquad Y_{k+1} = Y_k + hf(x_k, Y_k), \qquad Y_0 = y_0.$$

Here $x_k = x_0 + kh$. This is an example of a *Single-Step Method* which we write as

(II.5) $$Y_{k+1} = Y_k + h\phi(x_k, Y_k; h).$$

From the form of (II.5) one might think that

$$\phi(x_k, Y_k; h) = f(x_k, Y_k)$$

is the only "natural" choice. However, let me point out that the familiar Runge-Kutta method is also of this form, but with the more general form of ϕ.

THEOREM 1. *Let $f(x,y)$ be continuous in (x,y) and satisfy the Lipschitz condition* (II.3). *Moreover, let $\phi(x,y;h)$ also satisfy a Lipschitz condition in y. Then*

(II.6) $$\lim_{h\to 0+} \phi(x,y,h) = f(x,y)$$

is a necessary and sufficient condition for the convergence of the solution $\{Y_k\}$ of (II.5) to $y(\bar{x})$ in the limit as

$$h \to 0, \qquad k \to \infty, \qquad kh + x_0 \to x.$$

THEOREM 2. *Let $f(x,y)$ be continuous in (x,y) and satisfy the Lipschitz condition* (II.3); *let $\phi(x,y;h)$ also satisfy a Lipschitz condition. Moreover, let the "consistency" condition* (II.6) *be satisfied. Finally, let the truncation error be $O(h^p), p > 0$, i.e., if $y(x)$ is the solution of* (II.1), *then (if we have taken $x_0 = 0$ by a translation of the x-axis)*

$$y[(k+1)h] = y(kh) + h[\phi(x_k, y(kh); h) + O(h^p)].$$

Then, as $h \to 0$ and $kh \to \bar{x}$, we have

(II.7) $$\| Y_k - y(kh) \| = O(h^p).$$

We will omit the proof of Theorem 1, as it is technically complicated. However, let us give a proof of Theorem 2.

PROOF OF THEOREM 2. Let

$$E_k = Y_k - y(kh).$$

Then, from (II.5) and (II.7) we have

$$E_{k+1} = E_k + h\{[\phi(x_k, Y_k; h) - \phi(x_k, y(kh); h)] + O(h^p)\}.$$

Therefore, since ϕ also satisfies a Lipschitz condition,

$$\| E_{k+1} \| \leqq \| E_k \| + hL\{\| E_k \| + O(h^p)\},$$

i.e.,

$$\|E_{k+1}\| \leqq (1 + hL)\,\|E_k\| + Mh^{p+1},$$

or

$$\|E_{k+1}\| \leqq (1 + hL)^2\,\|E_{k-1}\| + [(1 + hL) + 1]Mh^{p+1},$$

or

$$\|E_{k+1}\| \leqq (1 + hL)^{k+1}\,\|E_0\| + \left[\sum_{j=0}^{k} (1 + hL)^j\right] Mh^{p+1}.$$

We sum the geometric progression and find that

$$\|E_{k+1}\| \leqq (1 + hL)^{k+1}\,\|E_0\| + \frac{(1 + hL)^{k+1} - 1}{1 + hL - 1}\, Mh^{p+1}.$$

That is

$$\|E_{k+1}\| \leqq e^{hL(k+1)}\left[\|E_0\| + \frac{M}{L}\, h^p\right].$$

If we now assume that $E_0 = 0$, we obtain the desired result.

Having these two theorems, we will leave the topic of single-step methods for the initial-value problem.

Of course, there are other methods of treating the initial-value problem. Let us consider the *Linear Multi-Step Methods*. The simplest such method is

$$Y_{k+1} = Y_{k-1} + 2hf(x_k, Y_k). \tag{II.8}$$

Notice that in this case we must specify both Y_0 and Y_1. Now Y_0 can be taken as y_0, but it is difficult to specify Y_1 exactly.

We now consider only the scalar case, i.e., $y(x)$ is a scalar, not a vector.

In general, we have constants $\alpha_0, \alpha_1, \cdots, \alpha_k, \beta_0, \cdots, \beta_k$, and we use the recurrence relation

$$\begin{aligned}\alpha_k Y_{n+k} + \alpha_{k-1} Y_{n+k-1} + \cdots + \alpha_0 Y_n \\ = h\{\beta_k f(x_{k+n}, Y_{k+n}) + \cdots + \beta_0 f(x_n, Y_n)\},\end{aligned} \tag{II.9}$$

or

$$\sum_{j=0}^{k} \alpha_j Y_{n+j} = h \sum_{j=0}^{k} \beta_j f(x_{n+j}, Y_{n+j}). \tag{II.9a}$$

Of course, we assume $\alpha_k \neq 0$. If $\beta_k = 0$, we say that (II.9) is an "explicit" linear multi-step method. On the other hand, if $\beta_k \neq 0$, then we have an "implicit" method.

EXAMPLE 3. Consider the linear multi-step method

$$\text{(II.10)} \qquad Y_{n+3} + \frac{3}{2} Y_{n+2} - 3Y_{n+1} + \frac{1}{2} Y_n = 3hf(x_{n+2}, Y_{n+2}).$$

One can easily verify that—provided $f(x, y)$ is nice enough—

$$y(x_{n+3}) + \frac{3}{2} y(x_{n+2}) - 3y(x_{n+1}) + \frac{1}{2} y(x_n) = 3hf(x_n, y(x_n)) + O(h^4).$$

That is, (II.10) is a consistent approximation to (II.1) and with a small truncation error.

Consider the two problems

$$y' = -y, \qquad y(0) = 1,$$

$$y' = y, \qquad y(0) = 1,$$

in the range $0 \leqq x \leqq 1$ with $h = 0.01$. For those of you who have access to a computer, I recommend these problems. You will find them very instructive. They show that Theorem 2 which applies to the single-step method, does not apply to an arbitrary multi-step method.

In any case, a simple analysis—but one which is too lengthy to give here—shows that the solutions of (II.10) are unstable and do not converge to the solution of (II.1). Moreover, this is true even in the simplest cases.

The results for the general Linear Multi-Step Method are too complicated to prove in the short space of time we have here. However, they are easy enough to state. (See Henrici [1] for details.)

Let

$$\text{(II.11)} \qquad \rho(\zeta) = \sum_{j=0}^{k} \alpha_j \zeta^j.$$

We have a *Stability Condition*: For all ζ which are roots of $\rho(\zeta) = 0$, we must have

$$\text{(II.11a)} \qquad |\zeta| \leqq 1.$$

Moreover, if ζ is a multiple root of $\rho(\zeta) = 0$, we must have

$$\text{(II.11b)} \qquad |\zeta| < 1.$$

And, as before, we have a *Consistency Condition*: This condition—in words—merely says that the solutions of (II.1), i.e., the solutions of the differential equation "almost" satisfy the difference equation (II.9). It is rather easy to verify that a necessary condition for consistency is

$$\sum_{j=0}^{k} \alpha_j = 0,$$

$$\sum_{j=0}^{k} (j\alpha_j - \beta_j) = 0. \tag{II.12}$$

DEFINITION. A linear multi-step method given by two sets of coefficients $\{\alpha_j\}, \{\beta_j\}$ is called convergent if the error $E_n = |Y_n - y(x_n)| \to 0$ as $h \to 0$ and $n \to \infty$ in such a way that $x_n \to x$, provided only $E_j \to 0$ for $j = 0, 1, \cdots, k-1$ and the function $f(x, y)$ is continuous in (x, y) for all y and $|x - x_0| < b$ (for some $b > 0$) and also $f(x, y)$ satisfies a Lipschitz condition in y.

THEOREM 3. *The linear multi-step method given by the two sets of coefficients* $\{\alpha_j\}$, $\{\beta_j\}$ *is convergent if and only if both the stability condition* (II.11a), (II.11b) *and the consistency condition* (II.12) *are satisfied.*

THEOREM 4. *Suppose the linear multi-step method* (II.9) *is convergent. Let* $y(x)$ *be a solution of* (II.1). *Assume also that*

$$\sum_{j=0}^{k} \alpha_j y(x_{n+j}) = h \sum \beta_j f(x_{n+j}, y(x_{n+j})) + O(h^{p+1}).$$

Then

$$|E_m| = O(h^p).$$

Now, let us mention another approach to our basic problem. This approach is motivated by the fact that may good linear multi-step methods are implicit, i. e., $\beta_k \neq 0$. Therefore in general the solution of (II.9) is hard to compute. Thus, we are led to *Predictor-Corrector Methods* of the form

$$Y^*_{n+k} + \sum_{j=0}^{k-1} a_j Y_{n+j} = h \sum_{j=0}^{k-1} b_j f(x_{n+j}, Y_{n+j}),$$

$$\sum_{j=0}^{k} \alpha_j Y_{n+j} = h \left\{ \rho_k f(x_{n+k}, Y^*_{n+k}) + \sum_{j=0}^{k-1} \beta_j f \right\}, \quad \alpha_k \neq 0. \tag{II.13}$$

The general idea here is to use a high-order predictor formula and a "stable" corrector formula.

Before we leave these *initial-value problems,* a few remarks are in order.

The motivation for linear multi-step methods is clearly the desire to use more accurate formulae. However, one should note that these methods can lead to many complications. First of all, one must have accurate methods for more "starting" values than are implied by the problem. Also, there is the problem of stability. Finally, there is a whole host of problems associated with the slow decay of certain components of the error which have been introduced by the linear multi-step method itself. Once more, let me recommend the book by Henrici [1].

Now, let us say a few words about "boundary-value" problems. Consider the problem

$$\begin{aligned} -(py')' &= f(x), \qquad 0 \leqq x \leqq 1, \\ y(0) &= y(1) = 0, \end{aligned} \tag{II.14}$$

where the function $p(x) \geqq p_0 > 0$ is a differentiable function.

We take $h = 1/M$, where M is an integer. Let Y_j represent an approximation to $y(jh)$, and let $p_{j+1/2} = p[(j + 1/2)h]$. One could try to approximate (II.14) by

$$\begin{aligned} Y_0 &= Y_M = 0, \\ -p_{j-1/2}\,Y_{j-1} + (p_{j-1/2} + p_{j+1/2})\,Y_j - p_{j+1/2}\,Y_{j+1} &= h^2 f(x_j). \end{aligned} \tag{II.15}$$

Now we have two problems:

(1) Can we solve these equations for $Y_1, Y_2, \cdots, Y_{M-1}$?

(2) Assuming the answer to (1) is yes, does the error

$$E_k = |\,Y_k - y(kh)\,| \to 0 \quad \text{as } h \to 0,\ k \to \infty\,?$$

In both cases the answer is "yes"! Let us look at the first question.

Consider a general tridiagonal system of linear equations of the form $Y_0 = Y_M = 0$:

$$a_j\,Y_{j-1} + b_j\,Y_j + c_j\,Y_{j+1} = Q_j$$

where

$$b_j \geqq |a_j| + |c_j|. \tag{II.16}$$

Then, when we look at the straight-forward elimination procedure, we discover the following algorithm. Let

$$G_0 = F_0 = 0, \tag{II.17}$$

$$\left.\begin{aligned} D_k &= b_k + a_k G_{k-1}, && \text{(II.17a)}\\ G_k &= -c_k / D_k, && \text{(II.17b)}\\ F_k &= (Q_k - a_k F_{k-1}) / D_k. \end{aligned}\right\} \qquad k = 1, 2, \cdots, M-1.$$

Then

$$Y_{M-1} = F_{M-1} \tag{II.18a}$$

and for $j < M-1$, we have

$$Y_j = G_j Y_{j+1} + F_j. \tag{II.18b}$$

Thus, our Equation (II.15) can be solved rather easily. Morover, the condition (II.16) guarantees that this procedure is numerically stable.

As for the second question, if we multiply (II.15) by Y_j and sum on j, we have (taking $Y_s = 0$ for $s < 0$ and $s > M$)

$$\sum_{(j)} [\, Y_j \cdot p_{j-1/2}(Y_j - Y_{j-1}) + Y_j \cdot p_{j+1/2}(Y_j - Y_{j+1})\,] = \sum_{(j)} h^2 f_j Y_j,$$

or

$$\sum_j p_{j-1/2} \left[\frac{Y_j - Y_{j-1}}{h} \right]^2 = \sum f_j Y_j,$$

or

$$h \sum_{(j)} \left[\frac{Y_j - Y_{j-1}}{h} \right]^2 \leqq \frac{1}{p_0} (h \sum |f_j|^2)^{1/2} \cdot (h \sum |Y_j|^2)^{1/2}. \tag{II.19}$$

And, for any set Z_j; $j = 0, 1, 2, \cdots, M$, with $Z_0 = Z_m = 0$, we have

$$h \sum_{(j)} |Z_j|^2 \leqq \frac{h^2}{2(1 - \cos \pi h)} h \sum \left[\frac{Z_j - Z_{j-1}}{h} \right]^2.$$

This last result can be established by elementary matrix theory. Since

$$\frac{h^2}{2(1 - \cos \pi h)} \to \frac{1}{\pi^2} \text{ as } h \to 0,$$

we can claim the existence of a constant $K > 0$ so that

$$\left[h \sum_{(J)} |Z_j|^2\right] \leqq K^2 \left[h \sum_{(J)} \left(\frac{Z_j - Z_{j-1}}{h}\right)^2\right].$$

Thus, using (II.19), we have

$$(h \sum |Y_j|^2)^{1/2} \cdot \left(h \sum \left(\frac{Y_j - Y_{j-1}}{h}\right)^2\right)^{1/2}$$

$$\leqq \frac{K}{p_0} (h \sum |f_j|^2)^{1/2} \cdot (h \sum |Y_j|^2)^{1/2}$$

i.e.

$$\text{(II.20)} \qquad \left(h \sum \left(\frac{Y_j - Y_{j-1}}{h}\right)^2\right)^{1/2} \leqq \frac{K}{p_0} (h \sum |f_j|^2)^{1/2}.$$

Let $k > r$. Then

$$Y_k - Y_r = h \sum_{j=r+1}^{k} \left[\frac{Y_j - Y_{j-1}}{h}\right].$$

Therefore,

$$|Y_k - Y_r| \leqq h \sum_{j=r+1}^{k} \left|\frac{Y_j - Y_{-1}}{\cdot h}\right|$$

$$\leqq ((k-r)h)^{1/2} \cdot \left(h \sum_{(J)} \left(\frac{Y_j - Y_{j-1}}{h}\right)^2\right)^{1/2}.$$

That is, using (II.20),

$$\text{(II.21a)} \quad |Y(kh) - Y(rh)| \leqq |kh - rh|^{1/2} \cdot \frac{K}{p_0} (h \sum |f_j|^2)^{1/2}.$$

And if $r = 0$,

$$\text{(II.21b)} \qquad |Y_k| \leqq \frac{K}{p_0} (h \sum |f_j|^2)^{1/2}.$$

It is now an easy matter to prove the convergence of the $\{Y_k\}$ to the solution of the boundary-value problem. The simplest approach is merely to observe that the error E_k satisfies a similar difference equation. However, in this case, the right-hand side $f_j \to 0$ as $h \to 0$. Hence

$$h \sum |f_j|^2 \to 0 \text{ as } h \to 0$$

and the convergence follows from (II.21b).

III. **Partial differential equations.** Once more, let us consider the Initial-Value Problem. Consider the special case of a first-order linear system of the form

$$\frac{\partial U}{\partial t} = P(x,t;D)\,U,$$

(III.1)

$$U(x,0) = U_0(x).$$

Here, $x = (x_1, x_2, \cdots, x_n)$ and U is a vector $(U_1, U_2, \cdots, U_N)$ and $P(x, t; D)$ is a matrix polynomial in the $(\partial/\partial x)$ with coefficients depending on (x,t).

Let us look at a very simple special case—

$$\frac{\partial U}{\partial t} = -\frac{\partial U}{\partial x},$$

(III.2)

$$U(x,0) = f(x).$$

One can easily prove that the solution to this problem is

(III.2a) $$U(x,t) = f(x-t).$$

Indeed, to verify that (III.2a) is a solution (assuming that $f(x)$ is differentiable) is an exercise in Calculus.

Even though we know the solution of this problem, let us look at an approximation

$$\frac{\mathcal{U}(x,t+k) - \mathcal{U}(x,t)}{k} = -\frac{\mathcal{U}(x+h,t) - \mathcal{U}(x,t)}{h},$$

which reduces to

(III.3) $$\mathcal{U}(x,t+k) = (1+\lambda)\,\mathcal{U}(x,t) - \lambda\,\mathcal{U}(x+h,t)$$

where $\lambda = k/h$. Repeated application of (III.3) leads to a formula of the form

(III.4) $$\mathcal{U}(x,nk) = \sum_{j=0}^{n} a_j\,\mathcal{U}(x+jh,0) = \sum_{j=0}^{n} a_j f(x+jh).$$

The exact values of the coefficients a_j are inessential for our present

argument. The important fact is that the value of $\mathscr{U}(x,t)$ depends only on the value of $f(x)$ at points to the *right* of x. On the other hand, from (III.2a), we see that the solution of the differential equation depends on a value of $f(x)$ at a point to the *left* of x, namely $(x-t)$. It is now an easy matter to construct an initial function $f(x)$ which is very smooth and the solutions of (III.3) cannot possibly converge to $f(x-t)$. For example, let $f(x) > 0$ for $x < 0$ and $f(x) = 0$ for $x \geqq 0$. Then, we see from (III.4) that $\mathscr{U}(x, nk) = 0$ for all $x > 0$. On the other hand, $U(x, nk) > 0$ for all $x < nk$.

Consequently let us try another approach—

$$\frac{\mathscr{U}(x,t+k) - \mathscr{U}(x,t)}{k} = -\frac{\mathscr{U}(x,t) - \mathscr{U}(x-h,t)}{h},$$

which reduces to

$$\text{(III.5)} \qquad \mathscr{U}(x,t+k) = (1-\lambda)\,\mathscr{U}(x,t) + \lambda\,\mathscr{U}(x-h,t).$$

In this case, an argument very similar to the one we have just given shows that we must take $\lambda \leqq 1$.

These examples illustrate the general situation. As in the case of ordinary differential equations, it is not enough to have a consistent approximation to the differential equation. Moreover, the restrictions on the difference schemes are usually restrictions on ratios of the step-lengths in the different coordinate directions.

Let us look at another example, the heat equation

$$\text{(III.6)} \qquad \frac{\partial U}{\partial t} = \frac{\partial^2 U}{\partial x^2}, \qquad U(x,0) = f(x), t > 0, \text{ all } x.$$

We try the difference scheme

$$\frac{\mathscr{U}(x,t+k) - \mathscr{U}(x,t)}{k} = \frac{1}{h^2}\{\mathscr{U}(x-h,t) - 2\,\mathscr{U}(x,t) + \mathscr{U}(x+h,t)\}.$$

In this case, the necessary condition is

$$\text{(III.7)} \qquad \frac{k}{h^2} \leqq \frac{1}{2}.$$

This result is particularly interesting because, unlike our earlier results, there does not seem to be any obvious relationship between (III.7) and the analytical properties of the solution of (III.6).

Actually such relationships do exist: for instance, if (III.7) is satisfied, then

$$\sup_x |\, \mathcal{U}(x, nk)\,| \leqq \sup_x |\, \mathcal{U}(x, 0)\,|, \tag{III.8}$$

and the physically interesting solutions of (III.6) satisfies a similar estimate. On the other hand, as we shall see, there are convergent difference schemes for the heat equation which do not enjoy property (III.8). Let me put it this way: Since (III.7) implies (III.8), it is easy to prove that the solutions $\mathcal{U}(x, nk)$ of the difference scheme converge to $U(x,t)$, the solution of (III.6), provided that (III.7) holds. However, it is not apparent that (III.7) is a necessary condition for convergence. However, this is the case for the difference equation proposed above.

Let us return to our general problem. If we select an $h = (h_1, h_2, \cdots, h_n)$ and k then a finite difference equation should give us approximations to the solution $U(x,t)$ at the lattice points $(j_1 h_1, j_2 h_2, \cdots, j_n h_n, rk)$. Let $\mathcal{U}(r)$ denote the "vector"

$$\{\, \mathcal{U}(j_1 h_1, \cdots, j_n h_n, rk)\,\}_{j_1, j_2, \cdots, j_n = -\infty}^{\infty}$$

and let $\|\, \mathcal{U}(r)\,\|$ denote some norm on these "vectors." For example, we could have

$$\|\, \mathcal{U}(r)\,\| = \sup_{j_s h_s} |\, \mathcal{U}(j_1 h_1, j_2 h_2, \cdots, j_n h_n, rk)\,|, \tag{III.9a}$$

or

$$\|\, \mathcal{U}(r)\,\| = h_1 h_2, \cdots, h_n \sum_{j_s} |\, \mathcal{U}(j_1 h_1, j_2 h_2, \cdots, j_n h_n, rk)\,|^2, \tag{III.9b}$$

etc.

If B is a linear operator (i.e., an infinite matrix in this case) acting on these vectors, we define

$$\|B\| = \sup_{x \neq 0} \frac{\|Bx\|}{\|x\|}. \tag{III.10}$$

Suppose we have a finite-difference approximation to (III.1) of the form

$$\begin{aligned} \mathcal{U}(r+1) &= B(r)\, \mathcal{U}(r) \\ \mathcal{U}(j_1 h_1, j_2 h_2, \cdots, j_n h_n, 0) &= U_0(j_1 h_1, j_2 h_2, \cdots, j_n h_n). \end{aligned} \tag{III.11}$$

In (III.11) we should write $\mathscr{U}(r+1;\ h)$ and $B(r;\ h)$ since these operators and vectors will depend on the lattice h and the increment k, etc.

DEFINITION. We say that the family of operators $\{B(r;\ h)\}$ is *stable* in the interval $0 \leqq t \leqq T$ and in the $\|\ \|$ norm if there is a constant c, depending on T, such that

$$\|B(r;h)\,B(r-1;h)\cdots B(j+1;h)\,B(j;h)\| \leqq C \tag{III.12}$$

for all $r,\ j$ with

$$0 \leqq j \leqq r \leqq T/k. \tag{III.12a}$$

The basic convergence argument is based on this notion and a simple argument which we saw earlier in our discussion of single-step methods for ordinary differential equations.

THEOREM 5. *Suppose* (III.1) *has a solution* $U(x,t)$. *Let* $U(r)$ *be the "vector" determined by* $U(x,rk)$, *i.e.,*

$$U(r) = \{U(j_1h_1, \cdots, j_nh_n, rk)\}_{j_1,j_2,\cdots,j_n=-\infty}^{\infty}.$$

Assume that (III.11) *is a "consistent" approximation to* (III.1), *i.e.*

$$U(r+1) = B(r)\,U(r) + a(r), \tag{III.13}$$

where

$$\|a(r)\| = O(k^{1+p}), \quad p > 0. \tag{III.14}$$

Finally, assume that the family $\{B(r;h)\}$ *is stable for* $0 \leqq t \leqq T$ *in the* $\|\ \|$ *norm.*

Then, for all $0 \leqq t \leqq T < \infty$ *of the form* $t = rk$, *we have*

$$\|U(r) - \mathscr{U}(r)\| = O(Tk^p). \tag{III.15}$$

PROOF. Let $E(r) = U(r) - \mathscr{U}(r)$. Then, from (III.13) we have

$$\begin{aligned} E(r+1) &= B(r)\,E(r) + a(r) \\ &= B(r)\,B(r-1)\,E(r-1) + B(r)\,a(r-1) + a(r) \\ &= B(r)\,B(r-1)\cdots B(0)\,E(0) \\ &\quad + \sum_{j=2}^{r} [B(r)\,B(r-1)\cdots B(j)]\,a(j-1) + a(r). \end{aligned}$$

Since the family $\{B(r;h)\}$ is stable, and $E(0) = 0$, we have

$$\|E(r)\| \leqq CRO(k^{1+p}) = C(rk)O(k^p),$$

that is

$$\|E(r)\| \leqq CTO(k^p),$$

Thus we have shown that, under reasonable conditions, stability and consistency imply convergence. A natural question is—what about the converse? In the appropriate theoretical setup, the answer is that stability is in fact also necessary for convergence. Let me refer you to the excellent book by Richtmyer [4]. Of course, as a practical matter, stability is absolutely essential!

For general difference schemes, there is no obvious way to establish the stability or instability. However, for certain cases we do have methods of analysis; and these results lead to useful"rules of thumb" which may be applied in the more general cases.

Consider the case where $P(x,t;D)$ has constant coefficients. That is

$$\begin{aligned}\frac{\partial U}{\partial t} &= P(D)U \\ &= \sum_{l=0}^{R} \sum_{l_1+l_2+\cdots+l_n=l} A_{l_1 l_2 \cdots l_n} \frac{\partial^l}{\partial x_1^{l_1} \cdots x_n^{l_n}} U\end{aligned} \tag{III.16}$$

where the $A_{l_1 \cdots l_n}$ are constant matrices.

Moreover, let us assume that the difference equation (III.11) takes the form

$$\mathcal{U}(r+1) = B\,\mathcal{U}(r) \tag{III.17}$$

where B is a constant matrix. More specifically, we assume

$$\begin{aligned}&\mathcal{U}(j_1 h_1, \cdots, j_n h_n; rk + k) \\ &= \sum_{|l_1|+|l_2|+\cdots+|l_n| \leqq R_1} B_{l_1 l_2 \cdots l_n}\, \mathcal{U}[(j_1 + l_1)k_1, \cdots, (j_n + l_n)h_n; rk]\end{aligned} \tag{III.18}$$

where the $B_{l_1,\cdots,l_n}$ are constant matrices. Consider the matrix-valued function

$$\beta(\theta_1, \theta_2, \cdots, \theta_n) = \sum B_{l_1 l_2 \cdots l_n} \exp(i \sum l_j \theta_j). \tag{III.19}$$

A rather straightforward application of Fourier analysis, which can be done in several ways, leads to the following conclusion: Let the norm be chosen as in (III.9b). Then

$$\| B^r \| = \max_{|\theta_j| \leq \pi} \| \beta^r(\theta_1, \theta_2, \cdots, \theta_n) \|_F$$

where $\| \ \|_F$ represents the Euclidean finite-dimensional matrix norm of $\beta^r(\theta_1, \theta_2, \cdots, \theta_n)$.

Thus, our problem has been reduced to a finite-dimensional problem. This problem is still not trivial. In fact, it is sometimes rather messy. However, we do have a method of analysis.

Let us return to our earlier examples. Consider the Equations (III.2) and the difference equations suggested for its solution. For the first, if we use Equation (III.3), we find that

$$\beta(\theta) = (1 + \lambda) - \lambda e^{i\theta}$$

and

$$\beta(\pi) = (1 + 2\lambda),$$

$$\| \beta^r(\pi) \| = (1 + 2\lambda)^r \to \infty \qquad \text{as } r \to \infty, \quad rk \leq T$$

if λ is a constant. Thus the first method is unstable as well as nonconvergent.

In the other case we find that

$$\beta(\theta) = (1 - \lambda) + \lambda e^{-i\theta},$$

$$|\beta(\theta)|^2 = [1 - \lambda(1 - \cos\theta)]^2 + \lambda^2 \sin^2\theta$$

$$= 1 - 2\lambda(1 - \cos\theta) + \lambda^2(1 - 2\cos\theta + \cos^2\theta) + \lambda^2 \sin^2\theta$$

$$= 1 + 2\lambda(\lambda - 1)\,(1 - \cos\theta).$$

Since $1 - \cos\theta \geq 0$, $\max |\beta(\theta)| \leq 1$ if and only if $\lambda \leq 1$. And, in this case, $\max|\beta^r(\theta)| = \max|\beta(\theta)|^r$.

Turning now to the heat equation (III.6) and the related difference equation, we find that

$$\beta(\theta) = 1 - 2\,\frac{k}{h^2}\,(1 - \cos\theta).$$

Thus $\beta(\theta)$ is real and $\beta(\theta) \leq 1$, while $\beta(\theta) \geq -1$ if and only if (III.7) holds.

This analysis enables us to handle differential equations and difference equations with constant coefficients. But what about the general problem of differential equations and difference equations with variable coefficients? In the general case, we have the following rule of thumb: For each value (x_0, t_0), $0 \leq t_0 \leq T$, consider the dif-

ferential equation and difference equations with all coefficients evaluated at (x_0, t_0). These difference equations are of the form we have analyzed. And, if for (x_0, t_0) the corresponding difference equations are "stable," then they are also stable in the variable coefficient case.

The validity of the above rule of thumb has not been established in complete generality. However, there are some fairly general results justifying this procedure.

Before proceeding, let us point out that if, with a finite difference equation of the form (III.11), which we write for short

$$\mathcal{U}(r+1) = B\mathcal{U}(r),$$

we associate the norm

$$\| \mathcal{U}(r) \|^2 = h_1 \cdot h_2 \cdot h_3 \cdots h_n \sum | \mathcal{U}(\cdots, r) |^2,$$

then

$$\text{(III.20)} \qquad \|B\| = \sup \frac{\|BU\|}{\|U\|} = \max \|\beta(\theta_1, \cdots, \theta_n)\|_F,$$

where $\| \ \|_F$ is as before.

Let us now consider as a further example the wave equation

$$\text{(III.21)} \qquad \frac{\partial^2 U}{\partial t^2} = \frac{\partial^2 U}{\partial x^2}.$$

We try the difference scheme

$$\frac{\mathcal{U}_k^{n+1} - 2\mathcal{U}_k^n + \mathcal{U}_k^{n-1}}{(\Delta t)^2} = \frac{\mathcal{U}_{k-1}^n - 2\mathcal{U}_k^n + \mathcal{U}_{k+1}^n}{(\Delta x)^2},$$

where we write $\mathcal{U}(k\Delta x, n\Delta t) = \mathcal{U}_k^n$. We transform the difference equation to

$$\text{(III.22)} \qquad \mathcal{U}_k^{n+1} = 2\mathcal{U}_k^n - \mathcal{U}_k^{n-1} + \lambda^2(\mathcal{U}_{k-1}^n - 2\mathcal{U}_k^n + \mathcal{U}_{k+1}^n),$$

where we have put $\lambda = \Delta t / \Delta x$. We could use a geometric argument to establish stability criteria, since we already know that the solution of (III.21) is $U = f(x+t) + g(x-t)$. The domain of dependence argument tells us that for stability we must have $\lambda \leqq 1$. It is important to recognize, however, that (III.22) is not of the form (III.11). The vector $\mathcal{U}^{n+1}$ depends on both $\mathcal{U}^n$ and $\mathcal{U}^{n-1}$.

To avoid this difficulty, we write (III.22) as the system

$$\text{(III.23)} \qquad \begin{aligned} w_k^{n-1} &= v_k^n, \\ v_k^{n+1} &= \lambda^2(v_{k-1}^n + v_{k+1}^n) + 2(1-\lambda^2)v_k^n - w_k^n. \end{aligned}$$

The matrix $\beta(\theta)$ defined by (III.19) is given by

$$\text{(III.24)} \qquad \beta = \begin{bmatrix} 0 & 1 \\ -1 & 2(1-\lambda^2) + 2\lambda^2\cos\theta \end{bmatrix}.$$

It is possible to analyze $\|\beta^r(\theta)\|$, but we will not do it. Instead we will study the necessary condition for stability, i.e., the eigenvalues are less than or equal to one in absolute value. The eigenvalues μ must satisfy

$$\text{(III.25)} \qquad \mu^2 - 2[1 - \lambda^2(1-\cos\theta)]\mu + 1 = 0.$$

It follows immediately that the product of μ_+ and μ_- must equal 1; if the roots are real, then they are either $+1$ or -1, or one is larger than the other in absolute value. Complex roots can only be unity in magnitude. This leads to the condition

$$\{1 - \lambda^2(1-\cos\theta)\}^2 \leqq 1,$$

or

$$\text{(III.26)} \qquad -1 \leqq 1 - \lambda^2(1-\cos\theta) \leqq 1,$$

which is true if and only if $\lambda \leqq 1$.

Taking another approach, assume the solution of (III.22) is of the form

$$\text{(III.27)} \qquad \mathscr{U}_k^n = \mu^n e^{i(k\Delta x)\theta} = \mu^n e^{ik\theta} \qquad (\Delta x = 1).$$

Substituting into (III.22) gives

$$\mu^{n+1} - 2\mu^n + \mu^{n-1} = \lambda^2\mu^n[e^{i\theta} - 2 + e^{i\theta}].$$

Factoring μ^{n-1} gives

$$\text{(III.28)} \qquad \mu^2 - 2[1 + \lambda^2(\cos\theta - 1)]\mu + 1 = 0,$$

which is identical to (III.25).

IV. **Practical problems in partial differential equations.** Consider the heat equation (III.6), but ask that it be satisfied in

$$0 \leqq x \leqq 1, \qquad t > 0,$$

with conditions given

$$U(x,0) = f(x), \qquad 0 \leqq x \leqq 1,$$

$$U(0,t) = g(t), \tag{IV.1}$$

$$U(1,t) = h(t), \qquad t > 0.$$

In the notation of (III.22), write the following family of difference equations:

$$\frac{\mathcal{U}_k^{n+1} - \mathcal{U}_k^n}{\Delta t}$$

$$= \alpha \frac{\mathcal{U}_{k-1}^n - 2\,\mathcal{U}_k^n + \mathcal{U}_{k+1}^n}{(\Delta x)^2} + (1-\alpha) \frac{\mathcal{U}_{k-1}^{n+1} - 2\,\mathcal{U}_k^{n+1} + \mathcal{U}_{k+1}^{n+1}}{(\Delta x)^2},$$

where $\lambda \equiv \Delta t/(\Delta x)^2$ and $0 \leqq \alpha \leqq 1$. Rewriting, we obtain

$$-\lambda(1-\alpha)\,\mathcal{U}_{k-1}^{n+1} + [1 + 2(1-\alpha)\lambda]\,\mathcal{U}_k^{n+1} - \lambda(1-\alpha)\,\mathcal{U}_{k+1}^{n+1}$$

$$= \alpha\lambda[\mathcal{U}_{k-1}^n - 2\,\mathcal{U}_k^n + \mathcal{U}_{k+1}^n] + \mathcal{U}_k^n, \tag{IV.2}$$

which, when applied to a system of mesh points, yields a tri-diagonal system of linear equations in terms of the known boundary conditions, which can be solved for any α. Is the system stable?

As before, assume the solution of (IV.2) is

$$\mathcal{U}_k^n = \beta^n e^{ik\theta}.$$

We then obtain

$$\beta = \frac{1 + 2\alpha\lambda(\cos\theta - 1)}{1 + 2(1-\alpha)\lambda(1 - \cos\theta)}. \tag{IV.3}$$

It is clear that $\beta \leqq 1$ for all α, λ, θ. To meet the stability condition $\beta \geqq -1$, it is necessary and sufficient that

$$(2\alpha - 1)\lambda \leqq 1/2. \tag{IV.4}$$

Note that if $\alpha \leqq 1/2$, (IV.4) is no restriction on λ, since λ is always $\geqq 0$; if $\alpha > 1/2$. on the other hand, λ is restricted.

To study the stability and convergence of (IV.2), you must make a detailed study of the tri-diagonal matrix in an equation for the error, and this gives the same result as (IV.4).

We now have a one parameter family of finite difference equations,

and in particular, if we take $\alpha < 1/2$, the equations are unconditionally stable.

To prove the convergence of the method for $\alpha = 0$, consider

$$\mathcal{U}_k^{n+1} - \mathcal{U}_k^n = \lambda\left[\mathcal{U}_{k-1}^{n+1} - 2\mathcal{U}_k^{n+1} + \mathcal{U}_{k+1}^{n+1}\right] \tag{IV.5}$$

and study $\max_k \mathcal{U}_k^{n+1}$. Then either $\max_k \mathcal{U}_k^{n+1}$ occurs on the boundary, or $\max_k \mathcal{U}_k^n \geqq \max_k \mathcal{U}_k^{n+1}$. For, suppose $\max_k \mathcal{U}_k^{n+1}$ is not on the boundary. At the interior point where $\mathcal{U}_k^{n+1}$ is a maximum,

$$\mathcal{U}_{k-1}^{n+1} - 2\,\mathcal{U}_k^{n+1} + \mathcal{U}_{k+1}^{n+1} \leqq 0.$$

Then, since $\lambda \geqq 0$, $\mathcal{U}_k^{n+1} \leqq \mathcal{U}_k^n$. But this was the k for which $\mathcal{U}_k^{n+1} = \max_k \mathcal{U}_k^{n+1}$. Repeating the argument for $n-1, n-2, \cdots, 0$, we conclude

$$\begin{aligned} &\max|\mathcal{U}_k^n| \leqq \max\{|f|, |g|, |h|\}, \\ &0 \leqq nN, \qquad 0 \leqq t \leqq N\Delta t, \\ &0 \leqq k \leqq M, \quad 0 \leqq x \leqq 1, \end{aligned} \tag{IV.6}$$

which implies the stability of the method.

To see that this insures convergence, consider the error equation

$$E_k^{n+1} - E_k^n = \lambda\left[E_{k-1}^{n+1} - 2E_k^{n+1} + E_{k+1}^{n+1}\right] + \sigma_k^n, \tag{IV.7}$$

where $|\sigma_k^n| \leqq L(\delta t)^{1+p}$, $p > 0$, and $E_k^0 = 0$, $E_0^n = E_M^n = 0$. Let $\max_k E_k^{n+1} = E_j^{n+1}$. If $j \neq 0$, $j \neq M$, then

$$|E_j^{n+1}| \leqq |E_j^n| + |\sigma_k^n| \tag{IV.8}$$

by arguments similar to those used to prove (IV.6). Now it is straightforward to show by continued inequalities that

$$\max_k |E_k^{n+1}| \leqq \max_k |E_k^n| + L\Delta t^{1+p}, \tag{IV.9}$$

and thus convergence is assured.

As an example of a two-dimensional problem, consider the heat equation (see [5], [6], and [7])

$$\frac{\partial U}{\partial t} = \frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} \tag{IV.10}$$

for $t > 0$, and (x, y) in some region R. The initial and boundary conditions are written in a form analogous to (IV.1).

If we write $\mathcal{U}_{k,j}^n$ for $\mathcal{U}(k\Delta x, j\Delta y, n\Delta t)$, we get the difference equation

$$\text{(IV.11)} \qquad \frac{\mathcal{U}_{k,j}^{n+1} - \mathcal{U}_{k,j}^n}{\Delta t} = \frac{\mathcal{U}_{k-1,j} - 2\,\mathcal{U}_{k,j} + \mathcal{U}_{k+1,j}}{(\Delta x)^2} + \frac{\mathcal{U}_{k,j-1}^{n+2} - 2\,\mathcal{U}_{k,j}^{n+2} + \mathcal{U}_{k,j+1}^{n+2}}{(\Delta y)^2},$$

The superscripts on the right-hand side have been temporarily omitted. If we put the superscript $n+1$ on each term, we could prove a min-max principle similar to (IV.6) and the corresponding convergence results, but the resulting system of equations are difficult to solve. To avoid this, consider the case of (IV.11) written as

$$\text{(IV.12)} \qquad \frac{\mathcal{U}_{k,j}^{n+2} - \mathcal{U}_{k,j}^{n+1}}{\Delta t} = \frac{\mathcal{U}_{k-1,j}^{n+1} - 2\,\mathcal{U}_{k,j}^{n+1} + \mathcal{U}_{k+1,j}^{n+1}}{(\Delta x)^2} + \frac{\mathcal{U}_{k,j-1}^{n+2} - 2\,\mathcal{U}_{k,j}^{n+2} + \mathcal{U}_{k,j+1}^{n+2}}{(\Delta y)^2},$$

where the superscript changes are obvious. This gives a tri-diagonal system which must be inverted for each value of k. Not carefully the difference in (IV.11) written as (IV.12) or written as

$$\text{(IV.13)} \qquad \frac{\mathcal{U}_{k,j}^{n+1} - \mathcal{U}_{k,j}^n}{\Delta t} = \frac{\mathcal{U}_{k-1,j}^{n+1} - 2\,\mathcal{U}_{k,j}^{n+1} + \mathcal{U}_{k+1,j}^{n+1}}{(\Delta x)^2} + \frac{\mathcal{U}_{k,j-1}^n - 2\,\mathcal{U}_{k,j}^n + \mathcal{U}_{k,j+1}^n}{(\Delta y)^2}.$$

The use of (IV.13), followed by (IV.12), forms the well-known alternating direction method, which is stable and therefore converges, provided Δt is given the same value for every pair of steps. The proof of stability is similar to that used for (IV.2).

To illustrate a practical problem involving an elliptic equation, a boundary value problem, consider

$$\text{(IV.14)} \qquad \frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} = f(x,y)$$

in a region R with $U = g(x,y)$ on the boundary of R.

As a first step, we must pick Δx and Δy, i.e., set up a basic lattice in R.

There are various ways to handle an irregular boundary. We can

modify the boundary to fit the mesh, i.e., use only mesh points which are within the boundary and estimate the influence of the boundary on the nearest interior mesh point by interpolation. Alternately, we can modify our difference equation for interior points near the boundary. Let us assume that we take the former method. Then we have the following formulation: If $(k\Delta x, j\Delta y)$ is an interior point,

$$\Delta_h \mathcal{U} \equiv \frac{\mathcal{U}_{k+1,j} - 2\,\mathcal{U}_{k,j} + \mathcal{U}_{k-1,j}}{(\Delta x)^2} + \frac{\mathcal{U}_{k,j-1} - 2\,\mathcal{U}_{k,j} + \mathcal{U}_{k,j+1}}{(\Delta y)^2} = f_{k,j}. \tag{IV.15}$$

If $k\Delta x, j\Delta y$ is a boundary point, then $\mathcal{U}_{k,j}$ is known.

Most of the effort in research involving (IV.15) is in finding methods for solving the resulting linear systems. However, it is easy to prove convergence of (IV.15) as follows:

Observe that if $f_{k,j} \geqq 0$ for all k,j, then the maximum of $\mathcal{U}_{k,j}$ is assumed on the boundary. The proof is by contradiction. Suppose the maximum is interior. Then, if it is at the point k,j, the left-hand side of (IV.15) is less than zero. But $f(x,y) \geqq 0$. Then (IV.15) is satisfied only if both sides are zero. Extending consideration to the neighboring points, then to their neighboring points, etc., it is clear that the maximum of $\mathcal{U}_{k,j}$ cannot occur in the interior of R unless $f \equiv 0$ and $\mathcal{U}$ is constant, including the boundary points.

Similar results follow for $f_{k,j} \leqq 0$ in R. We conclude that $\min v_{k,j}$ must occur on the boundary. These results imply the existence of a solution of the Equations (IV.15), since they are a linear system of equations in as many unknowns for which either there is a solution, or there exists a nontrivial solution of the homogeneous equation. But there is no such nontrivial solution, in view of the above results.

Let

$$w_{k,j} = \frac{(k\Delta x)^2 + (j\Delta y)^2}{4}, \tag{IV.16}$$

which is defined everywhere. We find that

$$\Delta_h w = 1. \tag{IV.17}$$

Let $F = \max|f_{k,j}|$ and $\mathcal{U}$ be a solution of $\Delta_h \mathcal{U} = f$. Then

$$\Delta_h(wF - \mathcal{U}) \geqq 0, \tag{IV.18}$$

which implies that $\max(wF - \mathcal{U})$ occurs on the boundary. Hence

$$\max(\mathcal{U}_{k,j}) \leqq (r^2/4)F + \max_{\text{bdy}} |\mathcal{U}|,$$

where r is at least the radius of the smallest circle which encloses R. Similarly, we can show that $\Delta_h(\mathcal{U} - wF) \leqq 0$, and hence

$$\max|\mathcal{U}_{k,j}| \leqq Cr^2F + \max_{\text{bdy}} |\mathcal{U}|,$$

and convergence is assured by applying this estimate to the error.

V. **Inversion of a matrix.** In general we have large matrices to invert. How do we invert them? Let us write a typical matrix in a particular structural form.

Write the matrix $V = (v_{ij})$ as a compound vector:

$$\text{(V.1)} \qquad V = \begin{pmatrix} V_1 \\ V_2 \\ \cdot \\ \cdot \\ \cdot \\ V_n \end{pmatrix} \quad \text{where } V_j = \begin{pmatrix} v_{1j} \\ v_{2j} \\ \cdot \\ \cdot \\ \cdot \\ v_{ij} \end{pmatrix}$$

i.e., we arrange everything by lines. V_j is the vector of unknowns in the jth line of any of the typical linear systems discussed above. Then the problem takes the form

$$\text{(V.2)} \qquad AV = Y$$

where A is a matrix, V is the vector of unknowns (V.1), and Y is a vector of known values arising from the boundary conditions.

The matrix A has the following structure:

$$\text{(V.3)} \quad A = \begin{bmatrix} D_1 & F_1 & & & 0 & \\ E_2 & D_2 & F_2 & & & \\ & \cdot & \cdot & \cdot & & \\ & & \cdot & \cdot & \cdot & \\ & 0 & & \cdot & \cdot & F_{m-1} \\ & & & & E_m & D_m \end{bmatrix} = D + E + F.$$

It is a block tri-diagonal matrix, which can be written as the sum of

three matrices, D, E, and F, as in (V.3). The matrices D_j are again tri-diagonal, and therefore are particularly easy to invert.

An iterative method (see the excellent book [5] by Varga) for finding the solution of a problem written in the form (V.2) is to write

$$A = P - N,$$

assuming the P^{-1} is easily inverted and A is nonsingular. (V.2) becomes

$$PV = NV + Y, \tag{V.4}$$

and we obtain an iterative equation by placing superscripts as shown:

$$PV^{(\nu+1)} = NV^{(\nu)} + Y. \tag{V.5}$$

If we define the error $E^{(\nu)} = V^{(\nu)} - V$, (V.4) and (V.5) give

$$E^{(\nu+1)} = (P^{-1}N)^{\nu+1}E^{(0)}. \tag{V.6}$$

From this it follows that the iterative method will be convergent if $\max|\lambda| < 1$ for all λ which are eigenvalues of $P^{-1}N$, which is equivalent to: For all λ such that $\det\{\lambda P - N\} = 0$.

The value of λ, if either of these conditions is satisfied, enables us to estimate the "cost" of alternative methods of iteration. Suppose we have two iteration schemes: (a) $A = P_0 - N_0$ with error $E_0^{(\nu)}$ and $\lambda_0 = \max|\text{eigenvalue of } P_0^{-1}N_0|$ and (b) $A = P_1 - N_1$ with error $E_1^{(\nu)}$ and $\lambda_1 = \max|$ eigenvalue of $P_1^{-1}N_1|$. Then it can be shown that

$$\|E_k^{(\nu+1)}\| \sim \lambda_k^{\nu+1}\|E_k^{(0)}\|, \qquad k = 0 \text{ or } 1. \tag{V.7}$$

Taking the logarithm of (V.7),

$$\frac{\log\left(\dfrac{\|E_k^{(\nu+1)}\|}{\|E_k^{(0)}\|}\right)}{\log\lambda_k} \sim \nu + 1,$$

which tells us how many iterations it would take to accomplish a fixed-ratio decrease in the norm of the error. It is therefore important to be able to estimate the value of λ.

To apply iteration to our problem, write $P = D$ and $N = -(E + F)$, and we obtain what is known as the Jacobi block iteration method:

$$DV^{(\nu+1)} = -(E+F)V^{(\nu)} + Y. \tag{V.9}$$

As an alternative but closely related method, consider the Successive Over-Relaxation method—SOR. Here

$$P = \frac{1}{\omega}(D + \omega E),$$

$$N = \frac{1}{\omega}[(1-\omega)D - \omega F],$$

where ω is some real number. This gives the iteration scheme

$$(D + \omega E)V^{(\nu+1)} = [(1-\omega)D + \omega F]V^{(\nu)} + \omega Y. \tag{V.10}$$

Since E is block lower triangular, we can invert $D + \omega E$ if we know the D_j^{-1}.

There is a relationship between the eigenvalues associated with each of the above methods which is shown in the following:

THEOREM 6. *Let* $\lambda = \max |$ *eigenvalue of Jacobi method*$|$ *and let* $\mu = \max |$ *eigenvalue of* SOR *method. Then*

$$(\mu + \omega - 1)^2 = \lambda^2\omega^2\mu.$$

PROOF. λ arises as a root of

$$\begin{vmatrix} \lambda D_1 & F_1 & & & 0 \\ E_2 & \lambda D_2 & F_2 & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & F_{m-1} \\ & 0 & & E_m & \lambda D_m \end{vmatrix} = 0, \tag{V.11}$$

and μ is a root of

$$\begin{vmatrix} \frac{\mu+\omega-1}{\omega}D_1 & F_1 & & & 0 \\ \mu E_2 & \frac{\mu+\omega-1}{\omega}D_2 & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & F_{m-1} \\ & 0 & & \mu E_m & \frac{\mu+\omega-1}{\omega}D_m \end{vmatrix} = 0. \tag{V.12}$$

Forming the matrix

$$T_{(\alpha)} = \begin{vmatrix} 1 & & & & 0 \\ & \alpha & & & \\ & & \alpha^2 & & \\ & & & \ddots & \\ 0 & & & & \alpha^{m-1} \end{vmatrix}$$

and computing $|T_{(\alpha)}|Q|T_{(\alpha)}^{-1}|$, where Q is the determinant (V.12), we obtain

$$\begin{vmatrix} \frac{\mu+\omega-1}{\omega}D_1 & \alpha^{-1}F_1 & & & 0 \\ \mu\alpha E_2 & \frac{\mu+\omega-1}{\omega}D_1 & \alpha^{-1}F_2 & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \alpha^{-1}F_{m-1} \\ 0 & & & \mu\alpha E_m & \frac{\mu+\omega-1}{\omega}D_m \end{vmatrix},$$

which we call Q_α. Now

$$|\alpha I|Q_\alpha = \begin{vmatrix} \alpha\left(\frac{\mu+\omega-1}{\omega}\right)D_1 & F_1 & & & \\ \alpha^2\mu_2^E & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & F_{m-1} \\ & & & \alpha^2\mu E_m & \frac{\alpha(\mu+\omega-1)}{\omega}D_m \end{vmatrix}.$$

If we put $\alpha^2\mu = 1$ and compare this with (V.11), we have the result stated in the Theorem. There are two interesting cases to consider in the SOR Method.

Case (i). $\omega = 1$. Then $\mu = \lambda^2$, and the SOR method is seen to be superior to Jacobi because (a) there is a reduced computer storage requirement, and (b) this method is approximately twice as fast as Jacobi, in view of (V.8).

Case (ii). Let $\omega =$ optimum value $= \omega_b$. One can show that $1 < \omega_b < 2$.

For the finite difference equation for the Laplacian operator, A is positive definite, and D is positive definite. Thus the Jacobi method eigenvalue is given by

$$\lambda = \max_X \frac{|((E+F)X, X)|}{(DX, X)}.$$

To conclude, let us consider the alternating direction method, discussed earlier, as applied to the Laplace equation.

Let us define the matrices H and V so that

$$(Hu)_{k,j} = -\{u_{k-1,j} - 2u_{k,j} + u_{k+1,j}\},$$
$$(Vu)_{k,j} = -\{u_{k,j-1} - 2u_{k,j} + u_{k,j+1}\}.$$

Then Equation (IV.14) can be written in the form

$$(\theta_x H + \theta_y V)X = Y, \tag{V.13}$$

where

$$\theta_x = \frac{\Delta y^2}{2(\Delta x^2 + \Delta y^2)}, \qquad \theta_y = \frac{\Delta x^2}{2(\Delta x^2 + \Delta y^2)},$$

and X is a vector of unknowns. Writing (V.13) as $(H_1 + V_1)X = Y$, we see our aim is to invert the matrix $H_1 + V_1$, a positive definite matrix, where H_1 and V_1 are positive definite themselves. To do this, put

$$(H_1 + rI)U^{m+1/2} = (rI - V_1)U^m + Y,$$
$$(V_1 + rI)U^{m+1} = (rI - H_1)U^{m+1/2} + Y,$$

with $r > 0$. The true solution satisfies both of these equations for any value of r. The error satisfies

$$E^{m+1} = (V_1 + rI)^{-1}(rI - H_1)(H_1 + rI)^{-1}(rI - V_1)E^m \equiv WE^m.$$

The dominant eigenvalue of W can be estimated in two particular cases.

Case I. If the region R is a rectangle, H and V commute, and thus every eigenvalue of W is of the form

$$\frac{(r - v_k)}{(v_k - r)} \frac{(r - h_j)}{(h_j + r)},$$

which is always less than 1 in magnitude, since each factor is less than one.

Case II. H and V do not commute. Then take

$$(rI + V_1)W(rI + V_1)^{-1} = W_0.$$

This has the same eigenvalues as W. However

$$W_0 = [(rI - H_1)(rI + H_1)^{-1}][(rI - V_1)(rI + V_1)^{-1}].$$

The terms in both brackets commute separately. Then

$$\|w_0\| \leqq \|(rI - H_1)(rI + H_1)^{-1}\| \, \|(rI - V_1)(rI + V_1)^{-1}\|.$$

Since the spectral radius of a symmetric matrix is its norm, and the spectral radius of any matrix is less than or equal to its $\| \ \|$ we conclude, as above: The method is convergent.

References

1. Peter Henrici, *Discrete variable methods in ordinary differential equations,* Wiley, New York, 1962.

2. Anonymous, *Modern computing methods,* Notes Appl. Sci., No. 16, National Physical Laboratory, London, 1957.

3. L. Fox, *The numerical solution of two-point boundary problems in ordinary differential equations,* Oxford Univ. Press, Oxford, 1957.

4. R. D. Richtmeyer, *Difference methods for initial-value problems,* Interscience, New York, 1957.

5. R. S. Varga, *Matrix iterative analysis,* Prentice-Hall, Englewood Cliffs, New Jersey, 1962.

6. J. Douglas, Jr., *Survey of numerical methods for parabolic differential equations,* Advances in Computers, Vol. 11, Academic Press, New York, 1961.

7. D. M. Young and T. G. Frank, *A survey of computer methods for solving elliptic and parabolic partial differential equations,* I. C. C. Bulletin, Vol. 2, No. 1, January, 1963.

8. A. A. Bennett, W. E. Milne and H. Bateman, *Numerical integration of differential equations,* Dover, New York, 1956.

UNIVERSITY OF WISCONSIN

J. Kevorkian

The Two Variable Expansion Procedure for the Approximate Solution of Certain Nonlinear Differential Equations

I. **Introduction.** This chapter is a revised version of the author's (1962) lecture notes([1]) which concerned an expansion procedure using two time scales for the asymptotic solution valid for long times of oscillations subject to small nonlinear forces.

A brief account of this two variable expansion procedure was also given in [5]. It should be pointed out that the material in the author's (1961) thesis [13], which is the basis of all the above publications, reflected to a great extent original ideas and suggestions by J. D. Cole.

In the study of oscillations for which a linear restoring force is perturbed by small nonlinear forces, one usually encounters two distinct groupings of the constants appearing in the problem into combinations with dimensions of time.

For example, for the case where the small perturbation is a damping, one time scale measures the relatively small period of the harmonic oscillations which are produced for zero damping, while the other scale determines the period over which the effect of damping becomes appreciable.

([1]) These are reported in [14]. The work in [13] and [14] was sponsored by the U.S. Air Force under Grant No. AF-AFOSR-62-256 and by the Douglas Aircraft Co., Inc., under Independent Research and Development Fund No. 81225-274 51953.

In the expansion procedure presented here two dimensionless time-like variables whose ratio is the small parameter of the problem are introduced explicitly in the calculations. If one treats these two times as distinct independent variables one can transform the initial value problem for an ordinary differential equation to one involving a partial differential equation in two times. The asymptotic solution of the transformed problem involves certain indeterminate functions which are then defined by postulating that the problem possesses a consistent asymptotic expansion which is uniformly valid for times of the order of the reciprocal of the small parameter.

In [13], the close relationship of the two variable method to the method of averaging of Krylov and Bogoliubov [16], and the technique discussed by Kuzmak [17], was pointed out. In [14] the more recent work of Struble [25], was also mentioned as being related. It is difficult to ascertain the precise origin of the idea of introducing two time variables (either explicitly as done here, or implicitly as in the method of averaging) in the development of an approximate solution. This concept which is implicit at least in all the work cited above, also appears in more recent publications dealing with the same problem as well as in other contexts.

The significant recent contribution to the problem (in English) has been the work of Bogoliubov and Mitropolsky [2], in which the theoretical framework of the method of averaging is discussed in detail and the original ideas of Krylov and Bogoliubov [16], extended and modified.

The existence of such a theoretical background obviates in a sense the need for pursuing a similar aspect of the present work. What has been attempted instead, is to cast the two variable method into a framework which is a generalization of the procedure of Lindstedt for periodic solutions, as reported in [21]. In so doing, it has also been attempted to compare and contrast the proposed approach to the work of Kaplun and Lagerstrom [12], in singular perturbations.

The relation between the two variable expansion and the modified method of averaging is discussed by Morrison in [20]. He reviews both methods and gives explicit formulas showing the equivalence of the results of both methods up to the second approximation. Morrison contrasts the generality of the modified method of av-

eraging to the relative simplicity of the two variable expansion procedure which would lead one to expect the latter to be less comprehensive in its applicability. It will be indicated in these notes (although no general proof will be attempted) that once the proper choice of variables is made, the two variable expansion procedure is applicable even to those cases where no general results are available and cursory examination indicates a failure of the expansion procedure.

We first consider the autonomous initial value problem (writing $\dot{y}$ for dy/dt, $\ddot{y}$ for d^2y/dt^2, etc.)

$$\ddot{y} + y + \epsilon f(y, \dot{y}) = 0, \tag{I.1a}$$

$$y(0) = a, \tag{I.1b}$$

$$\dot{y}(0) = b. \tag{I.1c}$$

where $\epsilon \ll 1$, and derive an asymptotic representation of the solution to $O(\epsilon)$ uniformly valid for times of order ϵ^{-1}. These general results are then applied to several specific examples of perturbation functions of the type given in (I.1a). We next consider various examples where f may have a more general form than in (I.1a) for instance by involving the time or ϵ or both. These examples are chosen in order to show how, after some preliminary considerations, the procedure can be used for these more general cases. Finally, we investigate a class of problems in celestial mechanics, where the motion is dominated by a central force field with small perturbations. These problems are shown to be essentially equivalent to (I.1a) with a more general form for f, and two simple examples are worked out.

The method discussed in these notes has also been applied by Eckstein, Shi and Kevorkian in [6] and [7] to two quite general satellite problems. In fact, in the latter reference it is shown that one needs three time variables to describe the behavior of the solution adequately for long times.

Other authors, e.g., Benney in [1], and Bretherton in [3], have used similar methods in problems of fluid mechanics, and references to the application of multiple-scale expansions for problems in plasma dynamics are given by McCune in [18].

II. Bounded oscillations. Formulation of the problem.

a. *General Remarks.* Consider the initial value problem

$$\ddot{y} + y + \epsilon f(y, \dot{y}) = 0, \qquad t \geqq 0, \tag{II.1a}$$

$$y(0) = a, \tag{II.1b}$$

$$\dot{y}(0) = b. \tag{II.1c}$$

Where ϵ is a very small positive constant and a, b are constants independent of ϵ. There is no loss of generality in assuming that $b = 0$ in (II.1c) for the autonomous equation (II.1a) as this merely fixes the origin of time.

The problem (II.1) may be regarded as the dimensionless formulation of the equation of motion of a mechanical system, say, consisting of a mass undergoing almost harmonic oscillations. Linearity is destroyed by the presence of a small perturbative force which may depend in a general way upon the displacement y and velocity $\dot{y}$.

The applicability of the initial value problem (II.1) is by no means restricted to that of nonlinear oscillations of systems with one degree of freedom. Many problems of interest in celestial mechanics are characterized by the fact that the motion in question is influenced by a predominantly central force field with small perturbations. Such is the case, for example, for the motion of a satellite around a planet when the effects of a thin atmosphere, a distant sun, or a slightly nonspherical gravitational field are considered. It will be shown in §V that in terms of appropriate variables the above "almost Keplerian" problems are essentially[2] equivalent to the initial value problem (II.1).

To fix ideas, consider the following initial value problem describing the motion of a particle of mass m restrained by a linear spring and damped by a force proportional to the cube of the speed:

$$m\frac{d^2\bar{y}}{d\bar{t}^2} + k\bar{y} + l\left(\frac{d\bar{y}}{d\bar{t}}\right)^3 = 0, \tag{II.2a}$$

$$\bar{y}(0) = A, \tag{II.2b}$$

$$\left.\frac{d\bar{y}}{d\bar{t}}\right]_{\bar{t}=0} = 0. \tag{II.2c}$$

In the above m, k, l and A are dimensional constants and $\bar{y}$ and

[2] Actually, one is led to a slightly more general case than (II.1) where f may also depend on ϵ and t as well as other dependent variables (defined by subsidiary first order equations) which reduce to constants when $\epsilon = 0$. The ideas of the present method apply equally well for these cases.

$\bar{t}$ dimensional variables. A dimensional analysis of the four constants appearing in this problem shows that there are two independent time scales which we may choose as

(II.3a) $$T_1 = (m/k)^{1/2},$$

(II.3b) $$T_2 = m^2/klA^2.$$

Note that we can always form a third time scale T_3 by multiplying T_1 by the dimensionless ratio

(II.3c) $$\epsilon = T_1/T_2 = lA^2k^{1/2}m^{-3/2}.$$

This ratio measures the relative importance of the damping and spring forces and the case $\epsilon \ll 1$, which is of interest here, corresponds to "small damping."

The time scale T_1 is a measure of the period of the oscillatory behavior of the system produced by the spring, while T_2 measures the period after which the cumulative effects of damping become important. The existence, in the physical context of this problem, of these two time scales is a fundamental feature of the method proposed here and will be further elaborated subsequently in this section. For the present purposes let us choose A and T_1 as the basic length and time scales to define the following normalized variables:

(II.4a) $$t = \bar{t}/T_1,$$

(II.4b) $$y = \bar{y}/A$$

in terms of which the initial value problem (II.2) reduces to (cf. (II.1)):

(II.5a) $$\ddot{y} + y + \epsilon\dot{y}^3 = 0,$$

(II.5b) $$y(0) = 1,$$

(II.5c) $$\dot{y}(0) = 0.$$

From the physical content of (II.5) it is obvious that the solution $y(t, \epsilon)$ is a bounded function of time. In general, it is easy to set down sufficient conditions on the function f such that the solution of the problem (II.1) be bounded. For example, by considering the solution curves in the phase plane, (II.1a) reduces to

(II.6a) $$R\frac{dR}{dt} = -\epsilon Vf(y, V)$$

where

$$R = (y^2 + V^2)^{1/2}, \tag{II.6b}$$

$$V = dy/dt. \tag{II.6c}$$

Thus, a sufficient condition for y to be bounded is that there exist some $R_0 < \infty$ such that $dR/dt \leqq 0$ whenever $R > R_0$. Such a condition is easily tested for any given f using (II.6a), and in this section we will consider only functions f having this property.

b. *Initially Valid Expansions, (Limit Process Expansions)*. Let y denote the exact solution of (II.1). We wish to define an expression which approximates y to a degree of accuracy which depends upon the smallness of ϵ. This is accomplished by the asymptotic expansion of y in the limit as $\epsilon \to 0$ (cf. [**10**] for definitions).

Let $\{\xi_n(\epsilon)\}$ be an asymptotic sequence as $\epsilon \to 0$ with $\xi_0 = 1$. We say that

$$h^{(N)}(t, \epsilon) = \sum_{n=0}^{N} h_n(t)\xi_n(\epsilon) \tag{II.7}$$

is an "initially valid" asymptotic expansion of y with respect to the sequence $\{\xi_n\}$ if the h_n are derived from y by the successive application of the following limit process for each $N = 0,1,2\cdots$:

$$\lim_{\epsilon \to 0} \frac{[y(t, \epsilon) - h^{(N)}(t, \epsilon)]}{\xi_N(\epsilon)} = 0, \quad t \text{ fixed; } t < \infty. \tag{II.8}$$

For a given function $y(t, \epsilon)$ and a suitable given sequence $\{\xi_n\}$, the h_n are unique if they exist. However, here y is known only to the extent that it is the solution of the differential equation (II.1), and the sequence $\{\xi_n\}$ can be somewhat arbitrary.

It is plausible (and quite often true) that for physically meaningful differential equations the h_n are the solutions of differential equations obtained by the recursive application of the limit process (II.8) to the original *differential equation*. For lack of any precise conditions on problems where this interchange of limits is valid, we shall merely regard it as a very plausible conjecture.

is easy to verify that for equation (II.1) only the sequence will lead to nontrivial equations for the h_n, and these are

$$\ddot{h}_0 + h_0 = 0, \tag{II.9a}$$

$$\ddot{h}_1 + h_1 = -f(h_0, \dot{h}_0), \tag{II.9b}$$

(II.9c) $$\ddot{h}_2 + h_2 = -h_1 f_y(h_0, \dot{h}_0) - \dot{h}_1 f_{\dot{y}}(h_0, \dot{h}_0),$$

(where $f_y(y, \dot{y}) = \partial f(y, \dot{y})/\partial y$ and $f_{\dot{y}}(y, \dot{y}) = \partial f(y, \dot{y})/\partial \dot{y}$). The general term on the right-hand side of the equation for h_n cannot be given concisely, but can be computed for each value of n by expanding f in its Taylor series in the neighborhood of the arguments h_0, $\dot{h}_0$.

Equations (II.1b) and (II.1c) imply that the h_n satisfy the initial conditions

(II.10a) $$h_0(0) = a, \qquad h_n(0) = 0, \ n \neq 0,$$

(II.10b) $$\dot{h}_n(0) = 0.$$

(Note: $b = 0$.) The solution of (II.9a) satisfying the conditions (II.10) is $h_0 = a \cos t$. Since h_0 is periodic, we may represent the the right-hand side of (II.9b) by its Fourier series expansion.

(II.11a) $$-f(h_0, \dot{h}_0) = \frac{a_0}{2} + \sum_{n=1}^{\infty} [a_n \cos nt + b_n \sin nt]$$

where

(II.11b) $$a_n = -\frac{1}{\pi} \int_0^{2\pi} f(a \cos t, -a \sin t) \cos nt \, dt,$$

(II.11c) $$b_n = -\frac{1}{\pi} \int_0^{2\pi} f(a \cos t, -a \sin t) \sin nt \, dt.$$

In general the Fourier series (II.11a) will be finite. For example, this is the case if f is a polynomial in y and $\dot{y}$.

The solution of (II.9b) satisfying the appropriate initial condition is given below in terms of the known coefficients a_n and b_n, which will have been determined by (II.11);

(II.12) $$h_1 = \left[\frac{b_1}{2} - \sum_{n=2}^{\infty} \frac{n b_n}{1 - n^2}\right] \sin t - \left[\frac{a_0}{2} + \sum_{n=2}^{\infty} \frac{a_n}{1 - n^2}\right] \cos t$$
$$+ \sum_{n=2}^{\infty} \frac{1}{1 - n^2} (a_n \cos nt + b_n \sin nt)$$
$$+ \frac{t}{2} (a_1 \sin t - b_1 \cos t) + \frac{a_0}{2}.$$

In a similar manner all the h_n can be successively computed.

We immediately notice the presence of the "mixed-secular" terms $-(a_1/2)\,t\sin t$ and $-(b_1/2)\,t\cos t$ in (II.12). In order to show how these terms arise in the expansion procedure let us choose $f = 2\dot{y}$.

For this choice of f the first two terms in the expansion for h are

$$h^{(1)} = h_0 + \epsilon h_1 = a\cos t + \epsilon a(\sin t - t\cos t) \tag{II.13}$$

whereas the exact solution of (II.1) is

$$y = ae^{-\epsilon t}[\cos(1-\epsilon^2)^{1/2}t + \epsilon(1-\epsilon^2)^{-1/2}\sin(1-\epsilon^2)^{1/2}t]. \tag{II.14}$$

For this example we can directly verify that $h_0 + \epsilon h_1$ is the initially valid expansion of (II.14) to order ϵ, and that the mixed-secular term $-\epsilon at\cos t$ arises from the nonuniform representation of the term $ae^{-\epsilon t}\cos(1-\epsilon^2)^{1/2}t$ for large times. Without computing the explicit formulas for the other h_n, let us note that further mixed-secular terms will appear in each of the h_n, and it is easy to verify for this linear equation that these terms are contributed by the non-uniform expansion of the terms $\epsilon^{-\epsilon t}$, $\sin(1-\epsilon^2)^{1/2}t$ and $\cos(1-\epsilon^2)^{1/2}t$ of the exact solution.

Although the expansion (II.13) for h is not uniformly valid for large times, it is a valid representation of (II.14) for any finite time interval and is analogous to an "inner expansion" ([3]) for a boundary-value singular perturbation problem as defined by Kaplun and Lagerstrom in [12]. To point out this analogy, let us note that as $\epsilon \to 0$ for any fixed value of T, the difference between the exact solution and the asymptotic expansion tends to zero uniformly over the interval $0 \leqq t \leqq T$.

If we now transform the independent variable t to $\tilde{t} = \epsilon t$, the domain of validity of the asymptotic expansion maps into the triangular region of the ϵ, $\tilde{t}$ plane defined by $0 \leqq \tilde{t} \leqq \epsilon T$.

This is in exact analogy with the domain of validity of an inner expansion for a boundary value problem in which $\tilde{t}$ corresponds to the outer variable.

c. *The Nonexistence of a Limit Process Expansion Valid for Long Times.* In view of the analogy between the inner expansion of a

([3]) Here and in the ensuing discussion concepts and definitions from the theory of Kaplun and Lagerstrom in [12] for matched asymptotic expansions will often be invoked.

singular perturbation problem and our initially valid expansion, it is natural to ask whether there exists a corresponding strict analogy between an outer expansion as defined by Kaplun and Lagerstrom in [12] and an expansion which will, for our case, be valid for large times. Unfortunately, such an analogy does not exist, and the reason for this will be discussed in some detail as it provides a great deal of insight into what will be needed to approximate the solution over long intervals of time.

Superficially, it is easy to verify that there does not exist a limit process by which a set of meaningful limiting differential equations, valid for large times, can be derived from (II.1). As will be shown presently, this fact is a reflection of the structure of the exact solution.

Consider a typical term of the form

$$y = S(\epsilon t; \epsilon)\, P(t; \epsilon) \tag{II.15}$$

where S represents a bounded, slowly varying amplitude modulating a periodic function P whose frequency is $\omega = 1 + O(\epsilon)$. For example, for the linear case where $f = 2\dot{y}$, $S = e^{-\epsilon t}$ and $P = \cos(1 - \epsilon^2)^{1/2}t$, and in general we expect the solution of (II.1) to consist of various combinations of such terms.

We wish to approximate y asymptotically as $\epsilon \to 0$ uniformly for large times, where the extent of the domain of validity in t will be defined more precisely later on. It is clear that for no limit process expansion in which $\epsilon t \to 0$ as $\epsilon \to 0$ can one uniformly represent S for large times, since all such expansions are equivalent to the Taylor series of S in the neighborhood of the origin.

In fact, a necessary condition for the development of S in an expansion valid for large times is use of a limit process wherein $\tilde{t} = \epsilon t$ is held fixed as $\epsilon \to 0$. Thus, the following expansion

$$S = \sum_{n=0}^{N} S_n(\tilde{t})\epsilon^n + O(\epsilon^{N+1}) \tag{II.16a}$$

follows from a knowledge of S by the repeated application of the limit process (with $\tilde{t}$ fixed, $\tilde{t} \neq 0$)

$$\lim_{\epsilon \to 0} \frac{1}{\epsilon^N}\left[S(\epsilon t; \epsilon) - \sum_{n=0}^{N} S_n(\tilde{t})\epsilon^n \right] = 0 \tag{II.16b}$$

for each $N = 0, 1, 2, \cdots$.

Barring nonuniformities for small times, a most unlikely occurrence for the present problem, the representation (II.16a) for S is uniformly valid over the interval $0 \leqq t \leqq T(\epsilon)$ for some $T(\epsilon)$ where $\epsilon T(\epsilon) = O(1)$. For future reference we denote the above domain in t by $D(\epsilon^{-1})$.

If, in addition, (II.16b) leads to a set of S_n each of which is a bounded function of $\tilde{t}$, then (II.16a) is uniformly valid for all t independently of ϵ.

Unfortunately, $\tilde{t}$ or any slow variable which corresponds to $t \to \infty$ as $\epsilon \to 0$ is totally unsuitable for describing the long-term behavior of the periodic function P. For, consider the Fourier series of P

$$P(t; \epsilon) = \frac{c_0(\epsilon)}{2} + \sum_{n=1}^{\infty} [c_n(\epsilon) \cos n\omega(\epsilon) t + d_n(\epsilon) \sin n\omega(\epsilon) t] \tag{II.17}$$

where we assume that the frequency ω may be asymptotically developed as

$$\omega = 1 + \sum_{n=1}^{N} \omega_i \epsilon^i + O(\epsilon^{N+1}) \tag{II.18}$$

for each $N = 1, 2, \cdots$.

From (II.17) we see that if t is replaced by $\tilde{t}/\epsilon$ (or any other slow variable), the limit process in which $\tilde{t}$ is finite as $\epsilon \to 0$ (i.e., $t \to \infty$) does not exist for any of the trigonometric terms involved in (II.17).

In fact, the answer to the question of the uniform representation for large times of a periodic function such as P defined by a differential equation such as (II.1) for the special case in which $S =$ constant is given by the method of Lindstedt as presented in [21]. Briefly, this involves transforming the independent variable t to t^+ according to

$$t^+ = \left(1 + \sum_{i=1}^{M} \omega_i \epsilon^i + O(\epsilon^{M+1})\right) t \tag{II.19}$$

in (II.1), which is then solved recursively using a limit process whereby t^+ is now held fixed as $\epsilon \to 0$. The unknown coefficients

ω_i occurring in the "stretched" variable t^+ are determined by requiring each term in the asymptotic development of P to be periodic. This development is then asymptotically equal to the one obtained by applying the same limit process to the exact function defined by (II.17) with ω as expanded in (II.18).

More precisely, let the coefficients $c_n(\epsilon)$ and $d_n(\epsilon)$ in (II.17) have the following asymptotic expansions in powers of ϵ:

$$c_n(\epsilon) = \sum_{i=0}^{N} c_{ni}\epsilon^i + O(\epsilon^{N+1}), \tag{II.20a}$$

$$d_n(\epsilon) = \sum_{i=0}^{N} d_{ni}\epsilon^i + O(\epsilon^{N+1}), \tag{II.20b}$$

where the c_{ni} and d_{ni} are numbers independent of ϵ. Then the asymptotic expansion of P to any order ϵ^N as obtained by the method of Lindstedt will be

$$P(t;\epsilon) = \sum_{i=0}^{N} P_i(t^+)\epsilon^i + O(\epsilon^{N+1}) \tag{II.21a}$$

where the P_i are

$$P_i(t^+) = \frac{c_{0i}}{2} + \sum_{n=1}^{\infty} (c_{ni}\cos nt^+ + d_{ni}\sin nt^+) \tag{II.21b}$$

and t^+ is defined by (II.19) with $M = N + 1$.

We note incidentally that $\omega(\epsilon)$, which is not in general known a priori for a periodic function defined by an initial value problem, can be determined asymptotically only to one order higher than the expansion for P. This is a consequence of the conditions which define ω_n, namely that the P_{n+1} be periodic.

We shall next calculate the domain of validity in t for the expansion (II.21a) with the c_n, d_n and t^+ as defined by (II.20a), (II.20b) and (II.19) respectively. Let us compute R_N, the difference between the exact value of P and its expansion to order ϵ^N

$$R_N = P(t;\epsilon) - \sum_{n=0}^{N} P_n(t^+)\epsilon^n.$$

Use of the various definitions in (II.17), (II.20) and (II.21) in the above and some manipulation yields:

$$R_N = \frac{1}{2}\left(c_0(\epsilon) - \sum_{m=0}^{N} c_{0m}\epsilon^m\right)$$
$$+ \sum_{n=1}^{\infty}\left[c_n(\cos n\omega t - \cos nt^+) + \left(c_n(\epsilon) - \sum_{m=0}^{N} c_{nm}\epsilon^m\right)\cos nt^+\right.$$
$$\left. + d_n(\sin n\omega t - \sin nt^+) + \left(d_n(\epsilon) - \sum_{m=0}^{N} d_{nm}\epsilon^m\right)\sin nt^+\right].$$

Now since $\omega t - t^+ = O(\epsilon^{N+2}t)$ whenever (II.21a) is given to order ϵ^N, we conclude by using the following trigonometric identities

$$\cos n\omega t - \cos nt^+ = -2\sin\frac{n}{2}(\omega t + t^+)\sin\frac{n}{2}(\omega t - t^+),$$

$$\sin n\omega t - \sin nt^+ = 2\cos\frac{n}{2}(\omega t + t^+)\sin\frac{n}{2}(\omega t - t^+),$$

that

$$\frac{R_N}{\epsilon^N} = O(\epsilon) + O(\epsilon^2 t) \quad \text{as } \epsilon \to 0.$$

Since $R_N\epsilon^{-N}$ must be $O(\epsilon)$ in order for (II.21a) to be the asymptotic development of (II.17), we conclude that with the c_n, d_n and ω_n as defined here, (II.21a) is uniformly valid for t belonging to $D(\epsilon^{-1})$ also.

We have shown that mutually contradictory requirements dictate the choice of variables for the uniform development for long times for each of the two types of functions that would in general appear simultaneously in the solution of (II.1). This means that the asymptotic development of the solution cannot be derived by a limit process expansion in terms of either of these variables, and strongly suggests the need of an expansion procedure involving the two times t^+, and $\tilde{t}$ simultaneously.

d. *An Expansion in Terms of Two Times.* It is clear from the discussion in the preceding section that one should use the product of the asymptotic expansions for S and P as given in (II.16) and (II.21a) respectively in order to represent the function defined by (II.15).

The product of (II.16) and (II.21a) gives

$$y = \sum_{i=0}^{N}\left(\sum_{k=0}^{i} S_k(\tilde{t})\,P_{i-k}(t^+)\right)\epsilon^i + O(\epsilon^{N+1}). \tag{II.22}$$

The reader can easily verify that (II.22):

(i) is an asymptotic expansion of y as $\epsilon \to 0$ (cf. [**10**, pp. 11-13], uniformly valid for times belonging to at least $D(\epsilon^{-1})$.

(ii) is not a limit process expansion in the sense that it cannot be constructed by the repeated application of *one* limit process, even if y were given explicitly.

(iii) trivially contains the initially valid expansion of y in the sense that this expansion would be recovered from (II.22) by the repeated application of the limit process (II.8).

Actually, all the above statements apply equally well to the "composite expansion" in the theory of Kaplun and Lagerstrom in [**12**], with the remark that in the present context, the initially valid expansion is not as important or interesting as its counterpart, the inner expansion of a problem in singular perturbations.

The following assumption, which is basic to the expansion procedure proposed, is plausible in light of the preceding discussion on the roles of the two variables t^+ and $\tilde{t}$ and nature of the functions they each depict.

We assume that the solution of the initial value problem (II.1) possesses a development of the form

$$y = F(t^+, \tilde{t}; \epsilon) = \sum_{n=0}^{N} F^{(n)}(t^+, \tilde{t})\, \epsilon^n + O(\epsilon^{N+1}) \tag{II.23}$$

which is unique (cf. discussion below and in connection with the evaluation of ω_2) and can be determined from (II.1) by treating t^+ and $\tilde{t}$ formally as being distinct variables and requiring (II.23) to satisfy the conditions demonstrated for (II.22).

The uniqueness of (II.23) is in the asymptotic sense, i.e., to any order ϵ^N, F, when regarded as a function of t and ϵ, is unique up to terms of order ϵ^N. Furthermore, since the labelling of various groupings of t and ϵ is intrinsically arbitrary in the final form of F, uniqueness implies that the uniformity of F to any other ϵ^N is preserved independently of any rearrangement of terms that may be necessitated by such a relabelling. Consider, for example, a linear function of $\tilde{t}$ occurring in $F^{(n)}$ for some n. Such a function must be excluded because, if $\tilde{t}$ were relabelled as ϵt^+, the function in question would *exactly* become classifiable as a term of $F^{(n+1)}$, thereby destroying the uniformity of F to $O(\epsilon^{n+1})$ in $D(\epsilon^{-1})$.

No attempt will be made in setting down conditions on f in (II.1) under which the assumed structure of the solution is valid.

The purpose of this work is to discuss the purely formal expansion procedure one is led to under these assumptions and to illustrate the various ideas by means of examples.

We now introduce the notation

$$F_1^{(n)} = \frac{\partial F^{(n)}}{\partial t^+}, \qquad F_2^{(n)} = \frac{\partial F^{(n)}}{\partial \tilde{t}}, \tag{II.24a}$$

$$F_{11}^{(n)} = \frac{\partial^2 F^{(n)}}{\partial (t^+)^2}, \qquad F_{12}^{(n)} = F_{21}^{(n)} = \frac{\partial^2 F^{(n)}}{\partial t^+ \partial \tilde{t}}, \qquad F_{22}^{(n)} = \frac{\partial^2 F^{(n)}}{\partial \tilde{t}^2}, \tag{II.24b}$$

and henceforth regard t^+ and $\tilde{t}$ as being distinct independent variables. Thus, the typical terms of (II.1) will have the following expansions:

$$\dot{y} = F_1^{(0)} + \sum_{n=1}^{N} \left(F_1^{(n)} + F_2^{(n-1)} + \sum_{k=1}^{n} \omega_k F_1^{(n-k)} \right) \epsilon^n + O(\epsilon^{N+1}), \tag{II.25a}$$

$$\begin{aligned} \ddot{y} = {} & F_{11}^{(0)} + \epsilon \left(F_{11}^{(1)} + 2\omega_1 F_{11}^{(0)} + 2F_{12}^{(0)} \right) \\ & + \epsilon^2 \left\{ \sum_{n=0}^{N} \left[\sum_{r=0}^{n+2} \left(\sum_{k=0}^{r} \omega_k \omega_{r-k} \right) F_{11}^{(n-r-2)} \right. \right. \\ & \left. \left. + 2 \sum_{k=0}^{n+1} \omega_k F_{12}^{(n-k+1)} + F_{22}^{(n)} \right] \epsilon^n \right\} + O(\epsilon^{N+1}), \end{aligned} \tag{II.25b}$$

$$\omega_0 = 1, \tag{II.25c}$$

$$\begin{aligned} f(y, \dot{y}) &= f^{(0)} + \epsilon f^{(1)} + O(\epsilon^2) \\ &= f(F^{(0)}, F_1^{(0)}) + \epsilon \left[F^{(1)} f_y(F^{(0)}, F_1^{(0)}) \right. \\ &\quad \left. + (F_1^{(1)} + F_2^{(0)} + \omega_1 F_1^{(0)}) f_{\dot{y}}(F^{(0)}, F_1^{(0)}) \right] + O(\epsilon^2). \end{aligned} \tag{II.25d}$$

Although there is no concise form for the general term in (II.25d), one can compute this routinely by expanding f in its Taylor series in the neighborhood of $F^{(0)}$ and $F_1^{(0)}$.

Substitution of the expansions (II.25) into (II.1) leads to the following set of partial differential equations governing the first three terms in the expansion for F

$$F_{11}^{(0)} + F^{(0)} = 0, \tag{II.26a}$$

$$F_{11}^{(1)} + F^{(1)} = -2F_{12}^{(0)} - 2\omega_1 F_{11}^{(0)} - f^{(0)}, \tag{II.26b}$$

$$\begin{aligned} F_{11}^{(2)} + F^{(2)} = {} & -2F_{12}^{(1)} - (2\omega_2 + \omega_1^2) F_{11}^{(0)} - F_{22}^{(0)} \\ & - 2\omega_1 (F_{11}^{(1)} + F_{12}^{(0)}) - f^{(1)}. \end{aligned} \tag{II.26c}$$

Since (II.23) is the composite expansion of y it must contain h and hence satisfy the initial conditions (II.1b) and (II.1c). These imply that

(II.27a) $$F^{(0)}(0,0) = a; \qquad F^{(n)}(0,0) = 0, \ n \neq 0;$$

(II.27b)
$$F_1^{(0)}(0,0) = b;$$
$$F_1^{(n)}(0,0) = -\left(F_2^{(n-1)}(0,0) + \sum_{k=1}^{n} \omega_k F_1^{(n-k)}(0,0)\right), \ n \neq 0.$$

The general solution of (II.26a) is

(II.28) $$F^{(0)}(t^+, \tilde{t}) = A^{(0)}(\tilde{t}) \sin t^+ + B^{(0)}(\tilde{t}) \cos t^+$$

where $A^{(0)}$ and $B^{(0)}$ are as yet undefined functions of $\tilde{t}$.

In order to compute $F^{(1)}$ we replace $f^{(0)} = f(F^{(0)}, F_1^{(0)})$ by its Fourier series

(II.29) $$f^{(0)} = \frac{A_0}{2} + \sum_{n=1}^{\infty} (A_n \cos nt^+ + B_n \sin nt^+)$$

where according to the definition of the A_n and B_n

(II.30a) $$A_n(A^{(0)}, B^{(0)}) = \frac{1}{\pi} \int_0^{2\pi} f(A^{(0)} \sin t^+ + B^{(0)} \cos t^+, A^{(0)} \cos t^+ - B^{(0)} \sin t^+) \cos nt^+ dt^+,$$

(II.30b) $$B_n(A^{(0)}, B^{(0)}) = \frac{1}{\pi} \int_0^{2\pi} f(A^{(0)} \sin t^+ + B^{(0)} \cos t^+, A^{(0)} \cos t^+ - B^{(0)} \sin t^+) \sin nt^+ dt^+.$$

These depend explicitly only on $A^{(0)}$ and $B^{(0)}$.

If we now use the expression for $F^{(0)}$ given in (II.28) to compute $F_{12}^{(0)}$ and substitute this result together with (II.29) into the right-hand side of (II.26b) we obtain

(II.31)
$$F_{11}^{(1)} + F^{(1)} = -\left[2\frac{dA^{(0)}}{d\tilde{t}} - 2\omega_1 B^{(0)} + A_1(A^{(0)}, B^{(0)})\right] \cos t^+$$
$$+ \left[2\frac{dB^{(0)}}{d\tilde{t}} + 2\omega_1 A^{(0)} - B_1(A^{(0)}, B^{(0)})\right] \sin t^+$$
$$- \frac{A_0}{2} - \sum_{n=2}^{\infty} (A_n \cos nt^+ + B_n \sin nt^+).$$

The requirement that (II.23) be the uniformly valid development of a bounded function over the domain $D(\epsilon^{-1})$ forbids the occurrence of terms which grow any faster than $\epsilon^{n+1}t$ for example, in each of the $F^{(n)}$. Therefore, the first two bracketed terms on the right-hand side of (II.31) must vanish identically because they contribute to the solution the mixed-secular terms proportional to $t^+\sin t^+$ and $t^+\cos t^+$ in the domain $D(\epsilon^{-1})$.

This condition provides the following system of ordinary differential equations for $A^{(0)}$ and $B^{(0)}$:

$$\frac{dA^{(0)}}{d\tilde{t}} - \omega_1 B^{(0)} + \frac{1}{2} A_1(A^{(0)}, B^{(0)}) = 0, \tag{II.32a}$$

$$\frac{dB^{(0)}}{d\tilde{t}} + \omega_1 A^{(0)} - \frac{1}{2} B_1(A^{(0)}, B^{(0)}) = 0, \tag{II.32b}$$

with initial conditions

$$A^{(0)}(0) = b, \tag{II.33a}$$

$$B^{(0)}(0) = a, \tag{II.33b}$$

which follows from (II.27) and (II.28).

The solution of (II.32) will define $A^{(0)}$, and $B^{(0)}$ in terms of $\tilde{t}$, the initial parameters a and b and the constant ω_1, which so far appears to be arbitrary.

The role of ω_1 in equations (II.32) for $A^{(0)}$ and $B^{(0)}$ needs some further comments. Let us introduce the orthogonal transformation

$$A^{(0)}(\tilde{t}) = \overline{A}^{(0)}(\tilde{t})\cos\omega_1\tilde{t} + \overline{B}^{(0)}(\tilde{t})\sin\omega_1\tilde{t}, \tag{II.34a}$$

$$B^{(0)}(\tilde{t}) = -\overline{A}^{(0)}(\tilde{t})\sin\omega_1\tilde{t} + \overline{B}^{(0)}(\tilde{t})\cos\omega_1\tilde{t}, \tag{II.34b}$$

which corresponds to a rotation of the $A^{(0)}$, $B^{(0)}$ axes (or equivalently, the phase angle of the oscillations described by $F^{(0)}$) at a uniform rate of ω_1 with respect to $\tilde{t}$. By virtue of the structure of $A_1(A^{(0)}, B^{(0)})$ and $B_1(A^{(0)}, B^{(0)})$ as given in (II.30) we can immediately verify that $\overline{A}^{(0)}$ and $\overline{B}^{(0)}$ satisfy the following reduced equations which are free of ω_1

$$\frac{d\overline{A}^{(0)}}{d\tilde{t}} = -\frac{1}{2} A_1(\overline{A}^{(0)}, \overline{B}^{(0)}), \tag{II.35a}$$

$$\frac{d\overline{B}^{(0)}}{d\tilde{t}} = \frac{1}{2} B_1(\overline{A}^{(0)}, \overline{B}^{(0)}). \tag{II.35b}$$

Thus, $\overline{A}^{(0)}$ and $\overline{B}^{(0)}$ depend on $\tilde{t}$ and the initial conditions, but not on ω_1. Upon substitution of the expressions in (II.34) into (II.28) for $F^{(0)}$ we find that this quantity is also free of ω_1 if one rewrites t^+ and $\tilde{t}$ in terms of t. In fact, $F^{(0)}$ becomes

$$F^{(0)} = \overline{A}^{(0)}(\epsilon t; a, b)\sin\left[(1 + \epsilon\omega_1 + \epsilon^2\omega_2 + O(\epsilon^3))t - \epsilon\omega_1 t\right] + \overline{B}^{(0)}(\epsilon t; a, b)\cos\left[(1 + \epsilon\omega_1 + \epsilon^2\omega_2 + O(\epsilon^3))t - \epsilon\omega_1 t\right] \tag{II.36}$$

in which the cancellation of ω_1 in the arguments is exhibited. Thus, in (II.19) we need not have included ω_1 and shall henceforth set $\omega_1 = 0$ for simplicity.

This result could have been anticipated directly by noting that any given trigonometric function of t^+ (which also includes $\epsilon\omega_1$ in its definition) can always be split into products of terms with arguments $\epsilon\omega_1 t$ and $[1 + \epsilon^2\omega_2 + O(\epsilon^3)]t$ respectively. The latter should then be regarded as functions of a new t^+ while the terms depending on $\epsilon\omega_1 t$ should be treated as slowly varying functions of $\tilde{t}$ to be consistent.

Thus, the cancellation of ω_1 in the explicit representation for $F^{(0)}$ of (II.36) is a reflection of the fact that no unique allocation exists for a trigonometric term with an argument proportional to ϵt. Since there exists no intrinsic criterion for the evaluation of ω_1, such as through boundedness or uniformity considerations, we shall henceforth take $\omega_1 = 0$ for the sake of simplicity. Thus, $\overline{A}^{(0)} = A^{(0)}$, $\overline{B}^{(0)} = B^{(0)}$ and the solution of these functions can be computed from (II.35).

Returning now to the solution of $F^{(1)}$, one finds the following result after eliminating the first harmonics from the right-hand side of (II.31):

$$F^{(1)}(t^+, \tilde{t}) = A^{(1)}(\tilde{t})\sin t^+ + B^{(1)}(\tilde{t})\cos t^+ - \frac{A_0}{2} + \sum_{n=2}^{\infty} \frac{1}{n^2 - 1}\left[A_n \cos nt^+ + B_n \sin nt^+\right]. \tag{II.37}$$

Again $A^{(1)}$, and $B^{(1)}$ are unknown functions of $\tilde{t}$ to be determined by conditions on $F^{(2)}$. We note from (II.37) and (II.27) that $A^{(1)}$ and $B^{(1)}$ satisfy the following initial conditions

$$A^{(1)}(0) = \sum_{n=2}^{\infty} \frac{n}{1 - n^2} B_n(b, a) - \frac{B_1}{2}(b, a), \tag{II.38a}$$

$$B^{(1)}(0) = \frac{A_0(b,a)}{2} + \sum_{n=2}^{\infty} \frac{A_n(b,a)}{1-n^2}. \tag{II.38b}$$

In order to compute $F^{(2)}$ we replace $f^{(1)}$ by its Fourier series

$$f^{(1)} = \frac{C_0}{2} + \sum_{n=1}^{\infty} (C_n \cos nt^+ + D_n \sin nt^+). \tag{II.39}$$

For a given function $f(y,\dot{y})$, the coefficients C_n and D_n are explicit functions of $A^{(0)}$, $B^{(0)}$, $A^{(1)}$ and $B^{(1)}$ and can be calculated routinely using the definition for $f^{(1)}$ (cf. (II.25d)) and the usual formulas for the coefficients of a Fourier series (cf. (II.30)). Moreover, since $A^{(0)}$ and $B^{(0)}$ have been determined from the solution of (II.35), the C_n and D_n can be reduced to known functions of $A^{(1)}$, $B^{(1)}$, $\tilde{t}$ and the initial parameters a, b. There is very little to be gained in attempting to develop elaborate formulas for the C_n and D_n since for most practical calculations (e.g., when f is a polynomial in its arguments) the various Fourier series, and products thereof reduce to finite combinations of trigonometric functions which can be readily calculated.

After substituting for the various terms appearing on the right-hand side of (II.26), we obtain the following result

$$\begin{aligned} F_{11}^{(2)} + F^{(2)} = &- \left[2\frac{dA^{(1)}}{d\tilde{t}} + \frac{1}{2}\frac{dB_1}{d\tilde{t}} - 2\omega_2 B^{(0)} + C_1 \right] \cos t^+ \\ &+ \left[2\frac{dB^{(1)}}{d\tilde{t}} + \frac{1}{2}\frac{dA_1}{d\tilde{t}} + 2\omega_2 A^{(0)} - D_1 \right] \sin t^+ - \frac{C_0}{2} \\ &+ \sum_{n=2}^{\infty} \left\{ \left[\frac{2n\, dA_n/d\tilde{t}}{n^2-1} - D_n \right] \sin nt^+ \right. \\ &\left. - \left[\frac{2n\, dB_n/d\tilde{t}}{n^2-1} + C_n \right] \cos nt^+ \right\}. \end{aligned} \tag{II.40}$$

Clearly, the boundedness of $F^{(2)}$ for large t^+ requires that $A^{(1)}$ and $B^{(1)}$ satisfy the following equations

$$2\frac{dA^{(1)}}{d\tilde{t}} + \frac{1}{2}\frac{dB_1}{d\tilde{t}} - 2\omega_2 B^{(0)} + C_1 = 0, \tag{II.41a}$$

$$2\frac{dB^{(1)}}{d\tilde{t}} + \frac{1}{2}\frac{dA_1}{d\tilde{t}} + 2\omega_2 A^{(0)} - D_1 = 0. \tag{II.41b}$$

The solution of equations (II.41) subject to the initial conditions (II.38) will define $A^{(1)}$ and $B^{(1)}$ in terms of $\tilde{t}$, the initial parameters a, b and the unspecified constant ω_2.

Unlike ω_1, ω_2 cannot be regarded as arbitrary, in the sense that a trigonometric function with the argument $\epsilon^2\omega_2 t$ cannot be uniformly represented in terms of $\tilde{t}$ for t in $D(\epsilon^{-1})$. Moreover, the determination of the ω_n for $n \geqq 2$ cannot hinge on the requirement that the development (II.23) be bounded for t in $D(\epsilon^{-1})$ since this is automatically satisfied once the mixed secular terms are eliminated.

We recall that one of the underlying assumptions on which the development (II.23) is based involves the independence of the uniformity of the expansion upon the rearrangement of terms due to a change in the labelling of functions of t and ϵ. We shall show in the examples to be discussed in §III that this condition will be sufficient to determine ω_2. Although no explicit calculations will be given, the extension of the present expansion procedure to higher orders can be carried out routinely. One requires the mixed-secular terms with respect to t^+ to vanish to each order ϵ^n. This provides ordinary differential equations which together with the initial conditions define the $A^{(n-1)}$ and $B^{(n-1)}$ in terms of $\tilde{t}$, the initial parameters a, b and ω_n. One then determines ω_n by the above-mentioned condition upon the uniformity of the expansion.

III. **Bounded oscillations. Examples.**

a. *Linear Damping.* $f = 2\dot{y}$. We now resume discussion of the simple example which was considered in §IIb in order to show the failure of the initially valid expansion for large times.

With $f^{(0)} = 2F_1^{(0)}$, and $f^{(1)} = 2(F_1^{(1)} + F_2^{(0)})$, the various definitions (or some simple direct calculations) give the following values of the Fourier coefficients:

$$\text{(III.1a)} \qquad A_1 = 2A^{(0)}; \qquad B_1 = -2B^{(0)}; \qquad A_n = B_n = 0, \quad n \neq 1;$$

$$C_1 = 2(A^{(1)} - B^{(0)});$$

$$\text{(III.1b)} \qquad D_1 = -2(B^{(1)} + A^{(0)});$$

$$C_n = D_n = 0, \quad n \neq 1.$$

The solution of (II.35), with $b = 0$ is then

(III.2a) $$A^{(0)}(\tilde{t}) = 0,$$

(III.2b) $$B^{(0)}(\tilde{t}) = ae^{-\tilde{t}},$$

a result which is immediately confirmed by the effect of damping on a linear system (cf. (II.14)). To $O(1)$ the solution is given by (II.28) with $t^+ = [1 + O(\epsilon^2)]t$ and $A^{(0)}(\tilde{t})$ and $B^{(0)}(\tilde{t})$ as defined above in (III.2).

We can evaluate $A^{(1)}$ and $B^{(1)}$ by solving (II.41) with the results of (III.1) and (III.2) and subject to the initial conditions (II.38) which now become $A^{(1)}(0) = a$, $B^{(1)}(0) = 0$. This gives

(III.3a) $$A^{(1)}(\tilde{t}) = a\left[\tilde{t}\left(\omega_2 + \frac{1}{2}\right) + 1\right]e^{-\tilde{t}},$$

(III.3b) $$B^{(1)}(\tilde{t}) = 0.$$

Consider the term proportional to $\tilde{t}e^{-\tilde{t}}$ in (III.3). Since one may relabel $\tilde{t}e^{-\tilde{t}}$ as $\epsilon te^{-\tilde{t}}$, this term when multiplied by $\sin t^+$ (cf. definition of $F^{(1)}$ in (II.37)) would be relegated to $F^{(2)}$ where it would violate uniformity in $D(\epsilon^{-1})$. We have assumed that the expansion to any order ϵ^n is uniquely defined by the governing equations to order ϵ^{n+1}. In fact, $F^{(2)}$ is also determined at this stage, except for $A^{(2)}$ and $B^{(2)}$, and the term in question must be excluded by setting $\omega_2 = -1/2$. The asymptotic representation of the solution is now determined to $O(\epsilon)$ in the form

(III.4)
$$y = ae^{-\tilde{t}}\cos\left[1 - \frac{\epsilon^2}{2} + O(\epsilon^3)\right]t + \epsilon ae^{-\tilde{t}}\sin\left[1 - \frac{\epsilon^2}{2} + O(\epsilon^3)\right]t + O(\epsilon^2)$$

which is uniformly valid for t in $D(\epsilon^{-1})$ ([4]).

b. *Damping Proportional to the Cube of the Speed* $f = \dot{y}^3$. The physical context of this example, where the two time scales as well as ϵ were calculated in terms of the dimensional constants of the problem, was introduced in §IIa. We note that here $f^{(0)} = F_1^{(0)3}$, which can be used in (II.30) to derive the following values of the

([4]) Actually, since for this simple example the exponential decay of the solution if exactly represented by the two-variable expansion, uniformity of the expansion is assured for all times.

Fourier coefficients that are needed for the first order solution

$$\text{(III.5a)} \qquad A_1 = \frac{3}{4} A^{(0)}(A^{(0)^2} + B^{(0)^2}); \qquad A_3 = \frac{A^{(0)}}{4}(A^{(0)^2} - 3B^{(0)^2});$$

$$\text{(III.5b)} \qquad A_n = 0, \quad n \neq 1, 3;$$

$$\text{(III.5c)} \qquad B_1 = -\frac{3}{4} B^{(0)}(A^{(0)^2} + B^{(0)^2}); \qquad B_3 = \frac{B^{(0)}}{4}(B^{(0)^2} - 3A^{(0)^2});$$

$$\text{(III.5d)} \qquad B_n = 0, \quad n \neq 1, 3.$$

The solution of equations (II.35) with the initial conditions (II.33) for $A^{(0)}$ and $B^{(0)}$ is:

$$\text{(III.6a)} \qquad A^{(0)}(\tilde{t}) = 0,$$

$$\text{(III.6b)} \qquad B^{(0)}(\tilde{t}) = 2(3\tilde{t} + 4)^{-1/2}.$$

Therefore the solution to $O(1)$ is

$$\text{(III.7)} \qquad y = 2(3\tilde{t} + 4)^{-1/2}\cos(1 + O(\epsilon^2))t.$$

According to (II.37) and (III.6) we have:

$$\text{(III.8)} \qquad F^{(1)} = A^{(1)}(\tilde{t})\sin t^+ + B^{(1)}(\tilde{t})\cos t^+ + \frac{B^{(0)^3}}{32}\sin 3t^+.$$

Rather than computing the C_n and D_n according to the general form of $f^{(1)} = 3F_1^{(0)^2}(F_1^{(1)} + F_2^{(0)})$, use of (III.6), (III.7) and (III.8) in computing the actual value of $f^{(1)}$ leads to considerable simplification([5]). We also note that since we are not interested in carrying out the explicit solution of $F^{(2)}$, we need only compute C_1 and D_1 in order to evaluate $A^{(1)}$ and $B^{(1)}$ (cf. (II.40) and (II.41)). A straightforward calculation gives

$$\text{(III.9a)} \qquad C_1 = \frac{3}{4} B^{(0)^2}\left(A^{(1)} - \frac{15}{32} B^{(0)^3}\right),$$

$$\text{(III.9b)} \qquad D_1 = -\frac{9B^{(0)^2}}{4} B^{(1)},$$

in which case Equations (II.41) become

([5]) This is why no effort was made in deriving explicit formulae for the C_n and D_n for a general $f^{(1)}$ in §IId.

$$\text{(III.10a)} \quad 2\frac{dA^{(1)}}{d\tilde{t}} + \frac{3}{4}B^{(0)2}A^{(1)} + \frac{9}{128}B^{(0)5} - 2\omega_2 B^{(0)} = 0,$$

$$\text{(III.10b)} \quad 2\frac{dB^{(1)}}{d\tilde{t}} + \frac{9}{4}B^{(0)2}B^{(1)} = 0.$$

The initial values to be satisfied by $A^{(1)}$ and $B^{(1)}$ are according to (II.38)

$$\text{(III.11a)} \quad A^{(1)}(0) = \frac{9}{32},$$

$$\text{(III.11b)} \quad B^{(1)}(0) = 0.$$

The solution of the system (III.10) subject to (III.11) is

$$\text{(III.12a)} \quad A^{(1)}(\tilde{t}) = (3\tilde{t}+4)^{-1/2}\left[\frac{3}{8}(3\tilde{t}+4)^{-1} + 2\omega_2\tilde{t} + \frac{15}{32}\right],$$

$$\text{(III.12b)} \quad B^{(1)}(\tilde{t}) = 0.$$

Again, uniqueness and its implications require the exclusion of the term proportional to ω_2 in (III.12a) and we set $\omega_2 = 0$. Thus, the solution correct to $O(\epsilon)$ is

$$\begin{aligned} y = {} & 2(3\tilde{t}+4)^{-1/2}\cos[1+O(\epsilon^3)]t \\ & + \epsilon(3\tilde{t}+4)^{-1/2}\left\{\left[\frac{3}{8}(3\tilde{t}+4)^{-1} + \frac{15}{32}\right]\sin[1+O(\epsilon^3)]t \right. \\ & \left. + \frac{1}{4}(3\tilde{t}+4)^{-1}\sin 3[1+O(\epsilon^3)]t\right\} + O(\epsilon^2). \end{aligned} \quad \text{(III.13)}$$

c. *Linear Damping and Nonlinear Restoring Force.* In this case we set $f = \dot{y} - cy^3$, where c is an arbitrary constant independent of ϵ. The first two terms of the expansion of the perturbing function are

$$\text{(III.14a)} \quad f^{(0)} = F_1^{(0)} - cF^{(0)3},$$

$$\text{(III.14b)} \quad f^{(1)} = (F_1^{(1)} + F_2^{(0)}) - 3cF^{(0)2}F^{(1)},$$

and (III.14a) leads to the following values of the Fourier coefficients

$$\text{(III.15a)} \quad A_1 = A^{(0)} - \frac{3cB^{(0)}}{4}(A^{(0)2} + B^{(0)2}); \quad A_3 = \frac{cB^{(0)}}{4}(3A^{(0)2} - B^{(0)2});$$

(III.15b) $$A_n = 0, \quad n \neq 1, 3;$$

$$B_1 = -B^{(0)} - \frac{3cA^{(0)}}{4}(A^{(0)2} + B^{(0)2});$$

(III.15c)

$$B_3 = \frac{cA^{(0)}}{4}(A^{(0)2} - 3B^{(0)2});$$

(III.15d) $$B_n = 0, \quad n \neq 1, 3.$$

Use of the amplitude and phase defined by

(III.16a) $$A^{(0)} = \rho \sin\theta,$$

(III.16b) $$B^{(0)} = \rho \cos\theta,$$

simplifies Equations (II.35) considerably and one can directly compute the following solution([6]) for the case $a = 1$, $b = 0$:

(III.17a) $$\rho = e^{-\tilde{t}/2},$$

(III.17b) $$\theta = \frac{3c}{8}(1 - e^{-\tilde{t}}).$$

We note that to order unity the cubic term in the restoring force does not affect the amplitude (cf. (III.16a) and (III.2b)) but introduces a phase shift which tends to the constant value $3c/8$.

The calculation of the coefficients C_1 and D_1 is straightforward, although somewhat tedious, for this example and will not be discussed.

d. *Rayleigh's Equation* (*van der Pol's Equation*). As a final example consider Rayleigh's Equation for which $f = -\dot{y} + (\dot{y}^3/3)$. The transformation $w = \dot{y}$ in (II.1) for the above f leads, after differentiation, to

(III.18) $$\ddot{w} - \epsilon(1 - w^2)\dot{w} + w = 0$$

which is van der Pol's equation, although some authors (e.g. Stoker in [23]) denote the former equation as van der Pol's. There is no relative merit for one form or the other for the present class of problems and we shall use the formulation in terms of y. One then calculates the following Fourier coefficients of $f^{(0)} = -F_1^{(0)} + (1/3)F_1^{(0)3}$:

([6]) An error in this solution, which appeared in [13] and [14], was pointed out by J. A. Morrison.

(III.19a) $$A_1 = \frac{A^{(0)}}{4}(A^{(0)^2} + B^{(0)^2} - 4),$$

(III.19b) $$A_3 = \frac{A^{(0)}}{12}(A^{(0)^2} - 3B^{(0)^2}),$$

(III.19c) $$A_n = 0, \qquad n \neq 1, 3,$$

(III.19d) $$B_1 = -\frac{B^{(0)}}{4}(A^{(0)^2} + B^{(0)^2} - 4),$$

(III.19e) $$B_3 = \frac{B^{(0)}}{12}(B^{(0)^2} - 3A^{(0)^2}),$$

(III.19f) $$B_n = 0, \qquad n \neq 1, 3.$$

The solution of (II.35) with $b = 0$ is then easily computed to be

(III.20a) $$A^{(0)}(\tilde{t}) = 0,$$

(III.20b) $$B^{(0)}(\tilde{t}) = 2a\left[a^2 + (4 - a^2)e^{-\tilde{t}}\right]^{-1/2}.$$

We note if $a = 2$ then $F^{(0)}$ which is defined as

(III.21) $$F^{(0)} = 2a\left[a^2 + (4 - a^2)e^{-\tilde{t}}\right]^{-1/2}\cos t^+$$

reduces to the correct periodic solution, namely $2\cos t^+$. It is also noted that this periodic solution is stable in the strong sense that for any value of a, the solution tends to the limit cycle as $\tilde{t} \to \infty$.

According to (II.37) $F^{(1)}$ is given by

(III.22) $$F^{(1)} = A^{(1)}(\tilde{t})\sin t^+ + B^{(1)}(\tilde{t})\cos t^+ + \frac{B^{(0)^3}}{96}\sin 3t^+.$$

In this case

(III.23) $$f^{(1)} = -(F_1^{(1)} + F_2^{(0)}) + F_1^{(0)^2}(F_1^{(1)} + F_2^{(0)})$$

and one can derive the following values of C_1 and D_1 from the available information

(III.24a) $$C_1 = \frac{A^{(1)}}{4}(B^{(0)^2} - 4) - \frac{5B^{(0)^5}}{128} + \frac{B^{(0)^3}}{4} - \frac{B^{(0)}}{2},$$

(III.24b) $$D_1 = \frac{B^{(1)}}{4}(4 - 3B^{(0)^2}).$$

Use of Equations (III.20) simplifies the value of A_1 and B_1 and one obtains

$$\frac{1}{2}\frac{dB_1}{d\tilde{t}} = -\frac{1}{8}\frac{d}{d\tilde{t}}[B^{(0)}(B^{(0)2}-4)]$$

(III.25a)

$$= \frac{B^{(0)}}{64}(4-B^{(0)2})(4-3B^{(0)2}),$$

(III.25b) $$\frac{1}{2}\frac{dA_1}{d\tilde{t}} = 0.$$

Equations (II.41) then become

(III.26a) $$2\frac{dA^{(1)}}{d\tilde{t}} + \frac{A^{(1)}}{4}(B^{(0)2}-4) + \frac{B^{(0)5}}{128} - \frac{B^{(0)}}{4} - 2\omega_2 B^{(0)} = 0,$$

(III.26b) $$2\frac{dB^{(1)}}{d\tilde{t}} - B^{(1)} + \frac{3}{4}B^{(0)2}B^{(1)} = 0.$$

The initial conditions computed from (II.38) reduce to

(III.27a) $$A^{(1)}(0) = \frac{a(3a^2-16)}{32},$$

(III.27b) $$B^{(1)}(0) = 0.$$

The solution of (III.26) subject to (III.27) gives

$$A^{(1)}(\tilde{t}) = \frac{B^{(0)}}{8}\log\left(\frac{B^{(0)}}{a}\right) + \frac{B^{(0)}}{64}(B^{(0)2}+5a^2-32)$$

(III.28a)

$$+ B^{(0)}\left(\omega_2 + \frac{1}{16}\right)\tilde{t},$$

(III.28b) $$B^{(1)}(\tilde{t}) = 0.$$

It is clear that we must set $\omega_2 = -1/16$ in order to avoid an inconsistent term in $F^{(2)}$, and this requirement defines $F^{(1)}$ completely. We note again that (since for $a = 2$, $B^{(0)}(\tilde{t}) = 2$) this case contains the periodic solution, and that as $\tilde{t} \to \infty$ for arbitrary a, the solution approaches the limit cycle.

IV. **Miscellaneous general examples.** In this section we will apply the method developed for solving the initial value problem (II.1) to a variety of examples in which the perturbation function f will be allowed to depend also on t and ϵ.

For the sake of simplicity, the general discussion of §IId will

not be paralleled here for the case where $f = f(y, \dot{y}, t, \epsilon)$. In fact, in this case it is not possible to choose a slow variable a priori. Moreover, a particular set of dependent and independent variables which may arise naturally from the physical context of the problem, may not be suitable for the uniform development of the solution.

Each of the examples presented in this chapter will illustrate one aspect of the preliminary analysis that is required before the formalism of the previous sections can be applied.

a. *Oscillator with a Slowly Varying Frequency.* Consider the initial value problem

$$\ddot{y} + \mu^2(\tilde{t})\, y = 0, \qquad \mu \neq 0, \quad \epsilon \ll 1, \tag{IV.1a}$$

$$y(0) = a, \tag{IV.1b}$$

$$\dot{y}(0) = b. \tag{IV.1c}$$

The slowly varying function μ depends on $\tilde{t} = \epsilon t$ and has various interpretations depending upon the physical context of (IV.1).

In order to study (IV.1) with the help of the procedure developed in previous sections, introduce the transformation

$$\frac{dt^+}{dt} = \mu(\epsilon t) \tag{IV.2}$$

in terms of which (IV.1) becomes

$$\frac{d^2y}{d(t^+)^2} + \epsilon g(\tilde{t}) \frac{dy}{dt^+} + y = 0, \tag{IV.3a}$$

$$y(0) = a, \tag{IV.3b}$$

$$\left.\frac{dy}{dt^+}\right]_{t^+=0} = \frac{b}{\mu(0)}, \tag{IV.3c}$$

where

$$g(\tilde{t}) = \mu^{-2} \frac{d\mu}{d\tilde{t}}. \tag{IV.3d}$$

Equation (IV.2) is a generalization of (II.19) for the case of a variable frequency. Note that here t^+ is not a linear function of t. In fact, except when μ is constant, this function is not the "frequency" of the problem, as evidenced from (IV.3a).

We assume y has the following expansion in two variables

$$y = \sum_{n=0}^{N} F^{(n)}(t^+, \tilde{t})\,\epsilon^n + O(\epsilon^{N+1}). \tag{IV.4}$$

The equations governing $F^{(0)}$ and $F^{(1)}$ are:

$$F^{(0)}_{11} + F^{(0)} = 0, \tag{IV.5a}$$

$$F^{(1)}_{11} + F^{(1)} = -g(\tilde{t})\,F^{(0)}_1 - 2\frac{F^{(0)}_{12}}{\mu}. \tag{IV.5b}$$

If the solution of (IV.5a)

$$F^{(0)} = A^{(0)}(\tilde{t}) \sin t^+ + B^{(0)}(\tilde{t}) \cos t^+ \tag{IV.6}$$

is used to compute the right-hand side of (IV.5b), we obtain

$$\begin{aligned} &F^{(1)}_{11} + F^{(1)} \\ &= -\left[gA^{(0)} + \frac{2}{\mu}\frac{dA^{(0)}}{d\tilde{t}}\right]\sin t^+ + \left[gB^{(0)} + \frac{2}{\mu}\frac{dB^{(0)}}{d\tilde{t}}\right]\cos t^+. \end{aligned} \tag{IV.7}$$

Clearly, both bracketed terms in (IV.7) must vanish if $F^{(1)}$ is to be bounded for t^+ in $D(\epsilon^{-1})$.

The solution of the resulting two differential equations is:

$$A^{(0)}(\tilde{t}) = A^{(0)}(0)\left(\frac{\mu(0)}{\mu}\right)^{1/2}, \tag{IV.8a}$$

$$B^{(0)}(\tilde{t}) = B^{(0)}(0)\left(\frac{\mu(0)}{\mu}\right)^{1/2}. \tag{IV.8b}$$

The initial conditions imply that $A^{(0)}(0) = b/\mu(0)$, $B^{(0)}(0) = a$, hence $F^{(0)}$ is given by

$$F^{(0)} = \left(\frac{\mu(0)}{\mu(\tilde{t})}\right)^{1/2}\left(a\cos t^+ + \frac{b}{\mu(0)}\sin t^+\right). \tag{IV.9}$$

This example is most elegantly solved by the technique developed by Gardner in [11], which applies to a large class of Hamiltonian systems possessing adiabatic invariants. The first-order result of (IV.9) confirms, for example, the adiabatic invariance of E/μ to $O(1)$, where E is the nonconstant energy defined by

$$2E = \dot{y}^2 + \mu^2 y^2. \tag{IV.10}$$

b. *Beating Oscillations.* Consider the following linear mechanical system

$$m\frac{d^2\bar{y}}{d\bar{t}^2} + k\bar{y} = F_0\cos\alpha\bar{t}, \tag{IV.11a}$$

$$\bar{y}(0) = \frac{d\bar{y}}{d\bar{t}}\Big]_{\bar{t}=0} = 0 \tag{IV.11b}$$

where m, k, F_0 and α are constants.

We wish to study the solution of (IV.11) for the case where α is close to the natural frequency of the system $(k/m)^{1/2}$.

Clearly the appropriate nondimensional variables are:

$$y = \frac{\bar{y}}{(F_0/k)}, \quad \bar{t} = \frac{t}{(m/k)^{1/2}}, \quad \epsilon = \frac{(k/m)^{1/2} - \alpha}{(k/m)^{1/2}} \ll 1. \tag{IV.12}$$

The dimensionless form of the initial value problem (IV.11) is

$$\ddot{y} + y = \cos(1-\epsilon)t = \cos(t - \tilde{t}), \tag{IV.13a}$$

$$y(0) = \dot{y}(0) = 0. \tag{IV.13b}$$

We can of course compute the following exact solution of the above system:

$$y = \frac{1}{\epsilon(2-\epsilon)}\left[(\cos\epsilon t - 1)\cos t + \sin t \sin\epsilon t\right]. \tag{IV.14}$$

We note that $\lim_{\epsilon\to 0} y(t,\epsilon) = (t/2)\sin t$, namely resonant oscillations.

However, for any positive value of ϵ, the motion is bounded and represents a long period oscillation modulating a short period oscillation.

We should note that, since for t in $D(\epsilon^{-1})$, $y = O(1/\epsilon)$, the correct expansion for y must have the form:

$$y = \sum_{n=-1}^{N} F^{(n)}(t,\tilde{t})\,\epsilon^n + O(\epsilon^{N+1}). \tag{IV.15}$$

The reader can verify that a routine application of the two variable formalism leads to the correct development of (IV.14) for t in $D(\epsilon^{-1})$.

The only essential difference between this and the previous ex-

amples was the fact that ϵy rather than y possessed the usual two variable expansion.

c. *Mathieu's Equation.* The dimensionless form of Mathieu's Equation is (cf. [23]):

$$\ddot{y} + (\delta + \epsilon \cos t)\, y = 0. \tag{IV.16}$$

In [23], Stoker also discusses the stability of the solution of (IV.16) in the δ, ϵ plane. He shows that corresponding to transitional values of δ and ϵ from stability to instability, (IV.16) possesses a periodic solution with period 2π or 4π, and that the solution of (IV.16) with general initial values is unstable(7) for the transitional values of δ and ϵ.

Thus, by finding all the functions $\delta(\epsilon)$ for which (IV.16) admits a periodic solution one can define the boundaries of the regions of stability and instability in the δ, ϵ plane.

It is especially easy to compute the periodic solutions and their corresponding transitional curves $\delta(\epsilon)$ if ϵ is small, by the method of Lindstedt. To summarize the results given in [23] and [19], we note the following.

The transitional curves intersect the $\epsilon = 0$ axis at the critical points $\delta_c = n^2/4$, $n = 0, 1, 2, \cdots$. Through all these points pass two transitional curves $\delta^{(n)}$ and ${}^{(n)}\delta$, except at the origin which admits only one such curve ${}^{(0)}\delta$. The region to the left of the curve ${}^{(0)}\delta$ corresponds to unstable solutions, and as each of the transitional curves are crossed by increasing δ, the stability of the solution of (IV.16) reverses.

The asymptotic representation of the first four $\delta^{(n)}$ and ${}^{(n)}\delta$ correct to $O(\epsilon^4)$ is given below(8):

$${}^{(0)}\delta = -\frac{\epsilon^2}{2} + \frac{7}{32}\epsilon^4, \tag{IV.17a}$$

$$\delta^{(1)} = \frac{1}{4} + \frac{\epsilon}{2} - \frac{\epsilon^2}{8} - \frac{\epsilon^3}{32} - \frac{1}{768}\epsilon^4, \tag{IV.17b}$$

$${}^{(1)}\delta = \frac{1}{4} - \frac{\epsilon}{2} - \frac{\epsilon^2}{8} + \frac{\epsilon^3}{32} - \frac{1}{768}\epsilon^4, \tag{IV.17c}$$

(7) By stability we shall mean boundedness for all nonnegative values of t.

(8) For the expansion of all the $\delta^{(n)}$ and ${}^{(n)}\delta$ correct to $O(\epsilon^6)$, the reader is referred to p. 17 of [19].

(IV.17d) $$\delta^{(2)} = 1 - \frac{\epsilon^2}{12} + \frac{5}{6{,}912}\epsilon^4,$$

(IV.17e) $$^{(2)}\delta = 1 + \frac{5}{12}\epsilon^2 - \frac{763}{6{,}912}\epsilon^4,$$

(IV.17f) $$\delta^{(3)} = \frac{9}{4} + \frac{\epsilon^2}{16} + \frac{\epsilon^3}{32} + \frac{13}{10{,}240}\epsilon^4,$$

(IV.17g) $$^{(3)}\delta = \frac{9}{4} + \frac{\epsilon^2}{16} - \frac{\epsilon^3}{32} + \frac{13}{10{,}240}\epsilon^4.$$

In this section we will investigate the behavior of the solutions of (IV.16) in neighborhoods of the transitional curves emanating from the first three critical points $\delta_c = 0, 1/4, 1$. More precisely we will solve Mathieu's equation with

(IV.18) $$\delta = \delta_c + \sum_{n=1}^{N} \delta_n \epsilon^n + O(\epsilon^{N+1}), \qquad \delta_c = 0, \frac{1}{4}, 1$$

where the δ_n are arbitrary numbers. For each δ_c, (IV.18) describes an N parameter family of curves in the (δ, ϵ) plane which pass through δ_c.

We make the following additional assumption regarding the structure of the solutions of (IV.16). For each δ_c, there exists a unique function $\eta_\delta(\epsilon)$, depending upon δ_c and with the property $\eta_\delta(\epsilon) = o(1)$ as $\epsilon \to 0$, such that when δ is in an unstable region this instability is depicted only by the unboundedness of the solution with respect to $\eta_\delta t$ and not t.

Let us adopt the initial conditions

(IV.19a) $$y(0) = b,$$

(IV.19b) $$\dot{y}(0) = a,$$

and in the first instance study the solutions of (IV.16) in the neighborhood of the origin of the δ, ϵ plane.

For this purpose we set (cf. (IV.18))

(IV.20) $$\delta = \epsilon\delta_1 + \epsilon^2\delta_2 + O(\epsilon^3).$$

We will show by actually carrying out the expansion, that in this case $\eta_\delta(\epsilon) = \epsilon$. We expand y in terms of t and $\tilde{t} = \epsilon t$ thus:

(IV.21) $$y = \frac{1}{\epsilon} \sum_{n=0}^{N} F^{(n)}(t, \tilde{t})\epsilon^n + O(\epsilon^N).$$

The leading term in this expansion is of order ϵ^{-1} in order to accommodate the nonzero initial condition $\dot{y}(0) = a$.

The equations and the initial conditions obtained by substituting the expansions (IV.21) and (IV.20) into (IV.16) and (IV.19) are:

$$F_{11}^{(0)} = 0, \tag{IV.22a}$$

$$F_{11}^{(1)} = -2F_{12}^{(0)} - (\delta_1 + \cos t)\, F^{(0)}, \tag{IV.22b}$$

$$F_{11}^{(2)} = -2F_{12}^{(1)} - (\delta_1 + \cos t)\, F^{(1)} - \delta_2 F^{(0)} - F_{22}^{(0)}, \tag{IV.22c}$$

$$F_{11}^{(n)} = -2F_{12}^{(n-1)} - (\delta_1 + \cos t)\, F^{(n-1)} - F_{22}^{(n-2)} - \sum_{k=2}^{n} \delta_k F^{(n-k)}, \quad n \geqq 2, \tag{IV.22d}$$

$$F^{(1)}(0,0) = b, \qquad F^{(n)}(0,0) = 0, \quad n \neq 1, \tag{IV.23a}$$

$$F_1^{(1)}(0,0) = a - F_2^{(0)}(0,0),$$
$$F_1^{(n)}(0,0) = -F_2^{(n-1)}(0,0), \quad n \neq 1. \tag{IV.23b}$$

We note that the homogeneous solutions of all the $F^{(n)}$ are

$$F^{(n)}(t, \tilde{t}) = A^{(n)}(\tilde{t})\, t + B^{(n)}(\tilde{t}). \tag{IV.24}$$

In view of our assumption regarding the boundedness of the $F^{(n)}$ for large t we must set all the $A^{(n)}(\tilde{t}) = 0$.

The solution for $F^{(0)}$ is therefore

$$F^{(0)} = B^{(0)}(\tilde{t}). \tag{IV.25}$$

The initial condition (IV.23a) for $F^{(0)}$ requires that

$$B^{(0)}(0) = 0. \tag{IV.26}$$

If we use the fact that $F^{(0)}$ is a function of $\tilde{t}$ only in (IV.22b) we obtain

$$F_{11}^{(1)} = -\delta_1 B^{(0)}(\tilde{t}) - B^{(0)}(\tilde{t}) \cos t. \tag{IV.27}$$

Again, in order that $F^{(1)}$ be a bounded function of t, δ_1 must vanish, and this produces the correct stability criterion in the neighborhood of the origin. For, to $O(\epsilon)$, the transitional curve is indeed $\delta = 0$.

If we now solve for $F^{(1)}$ we obtain

$$F^{(1)} = B^{(1)}(\tilde{t}) + B^{(0)}(\tilde{t}) \cos t. \tag{IV.28}$$

The initial conditions for $F^{(1)}$ imply that

$$B^{(1)}(0) = b, \qquad \left.\frac{dB^{(0)}}{d\tilde{t}}\right]_{\tilde{t}=0} = a. \tag{IV.29}$$

If we now use the preceeding expressions for $F^{(0)}$ and $F^{(1)}$ to evaluate the right-hand side of (IV.22c) we obtain

$$F_{11}^{(2)} = 2\frac{dB^{(0)}}{d\tilde{t}}\sin t - \left[\frac{d^2B^{(0)}}{d\tilde{t}^2} + \left(\delta_2 + \frac{1}{2}\right)B^{(0)}\right] - B^{(1)}(\tilde{t})\cos t - \frac{B^{(0)}}{2}\cos 2t. \tag{IV.30}$$

The bracketed term on the right-hand side of (IV.30) is a function of t only, and hence must vanish in order that $F^{(2)}$ be bounded as $t \to \infty$.

The solution of the equation obtained by setting this bracketed term equal to zero is:

$$\text{(IV.31a)} \quad B^{(0)} = a\left(\delta_2 + \frac{1}{2}\right)^{-1/2} \sin\left(\delta_2 + \frac{1}{2}\right)^{1/2}\tilde{t}, \qquad \delta_2 \geqq -\frac{1}{2},$$

$$\text{(IV.31b)} \quad B^{(0)} = a\left(-\delta_2 - \frac{1}{2}\right)^{-1/2} \sinh\left(-\delta_2 - \frac{1}{2}\right)^{1/2}\tilde{t}, \qquad \delta_2 < -\frac{1}{2}.$$

Thus we have established the correct stability criterion for δ_2, and if we proceed with the higher order terms we will successively be able to compute more accurate expressions for the transitional curve and the solution on either side of this curve.

Let us now return to the question of the choice of $\eta_\delta(\epsilon)$. If we had chosen an $\eta_\delta \not\sim \epsilon$, either one of the following two eventualities would have arisen:

a. For $\epsilon = o(\eta_\delta)$ as $\epsilon \to 0$, it would have been impossible to make y a bounded function of $\eta_\delta t$ for values of δ_n consistent with the transitional curve. This can be seen easily if we consider a typical term in the expansion such as $\sin(\delta_2 + (1/2))^{1/2}t = \sin(\delta_2 + (1/2))^{1/2}\epsilon t$. If, for example, we have chosen $\eta_\delta(\epsilon) = \epsilon^{1/2}$ (i.e., $\tilde{t} = \epsilon^{1/2}t$) the term in question would have appeared in the solution in terms of its nonuniform expansion in powers of $\tilde{t}\epsilon^{1/2}$. This would have led to a term proportional to $\epsilon^{1/2}(\delta_2 + (1/2))^{1/2}\tilde{t}$ in the solution, and to the erroneous conclusion that $\delta_2 = -1/2$ was the *only* possible value for stability.

b. For $\eta_\delta(\epsilon) = o(\epsilon)$ as $\epsilon \to 0$, it would have been impossible to make y a bounded function of t for the following reason. If the two variables we are expanding in are t and, say $\tilde{t} = \epsilon^2 t$, the above term would appear in the solution in terms of its nonuniform expansion in powers of ϵt thus $(\delta_2 + (1/2))^{1/2}\epsilon t + \cdots$.

We conclude from the above that if we restrict ourselves to those solutions which lie on the family of curves to which the transitional curve belongs, (i.e., the m-parameter family with $m \leqq N$, obtained by varying the nonzero coefficients $\delta_1, \cdots, \delta_N$ appearing in the transitional curve), then we have a unique way of determining the second time variable in our expansion.

In the preceeding example, and in what follows the choice of $\tilde{t}$ was verified by actually carrying out the expansions for two other η_δ's in adjacent order classes. In all cases the results were as outlined above.

Next, let us consider the family of solutions in the neighborhood of $\delta_c = 1/4$. Here again the appropriate function $\eta_\delta(\epsilon)$ is ϵ, hence $\tilde{t} = \epsilon t$. We expand y in terms of t and $\tilde{t}$ in the form

$$y = \sum_{n=0}^{N} F^{(n)}(t, \tilde{t})\,\epsilon^n + O(\epsilon^{N+1}). \tag{IV.31}$$

The equations and initial conditions governing $F^{(0)}$ and $F^{(1)}$ are then easily calculated

$$F^{(0)}_{11} + \frac{F^{(0)}}{4} = 0, \tag{IV.32a}$$

$$F^{(1)}_{11} + \frac{F^{(1)}}{4} = -2F^{(0)}_{12} - \delta_1 F^{(0)} - F^{(0)}\cos t, \tag{IV.32b}$$

$$F^{(0)}(0,0) = b, \qquad F^{(1)}(0,0) = 0, \tag{IV.33a}$$

$$F^{(0)}_1(0,0) = a, \qquad F^{(1)}_1(0,0) = -F^{(0)}_2(0,0). \tag{IV.33b}$$

The solution of (IV.32a) for $F^{(0)}$ gives

$$F^{(0)} = A^{(0)}(\tilde{t})\sin\frac{t}{2} + B^{(0)}(\tilde{t})\cos\frac{t}{2}. \tag{IV.34}$$

The initial conditions imply that

$$B^{(0)}(0) = b, \qquad A^{(0)}(0) = 2a. \tag{IV.35}$$

When this expression for $F^{(0)}$ is substituted into (IV.32b), we obtain

$$F_{11}^{(1)} + \frac{F^{(1)}}{4} = \left[\frac{dB^{(0)}}{dt} + \left(\frac{1}{2} - \delta_1\right) A^{(0)}\right] \sin\frac{t}{2}$$

$$- \left[\frac{dA^{(0)}}{d\tilde{t}} + \left(\delta_1 + \frac{1}{2}\right) B^{(0)}\right] \cos\frac{t}{2} \tag{IV.36}$$

$$- \frac{A^{(0)}}{2} \sin\frac{3}{2}t - \frac{B^{(0)}}{2} \cos\frac{3}{2}t.$$

Thus, the boundedness of $F^{(1)}$ requires that $A^{(0)}$ and $B^{(0)}$ satisfy the equations

$$\frac{dA^{(0)}}{d\tilde{t}} = -\left(\delta_1 + \frac{1}{2}\right) B^{(0)}, \tag{IV.37a}$$

$$\frac{dB^{(0)}}{d\tilde{t}} = \left(\delta_1 - \frac{1}{2}\right) A^{(0)}. \tag{IV.37b}$$

The solution of (IV.37) falls into two categories. First if $\delta_1 > 1/2$ or $\delta_1 < -1/2$ the solutions are stable, and we have

$$A^{(0)} = \mp b \left(\frac{2\delta_1 + 1}{2\delta_1 - 1}\right) \sin\left(\delta_1^2 - \frac{1}{4}\right)^{1/2} \tilde{t} + 2a \cos\left(\delta_1^2 - \frac{1}{4}\right)^{1/2} \tilde{t}, \tag{IV.38a}$$

$$B^{(0)} = b \cos\left(\delta_1^2 - \frac{1}{4}\right)^{1/2} \tilde{t} \pm 2a \left(\frac{2\delta_1 - 1}{2\delta_1 + 1}\right)^{1/2} \sin\left(\delta_1^2 - \frac{1}{4}\right)^{1/2} \tilde{t}, \tag{IV.38b}$$

where the upper or lower signs are to be taken depending on whether $\delta_1 > 1/2$ or $\delta_1 < -1/2$ respectively.

Conversely, if $-1/2 < \delta_1 < 1/2$, the solutions are unstable, and we have

$$A^{(0)} = -b \frac{k_1}{k_2} \sinh(k_1 k_2 \tilde{t}) + 2a \cosh(k_1 k_2 \tilde{t}), \tag{IV.39a}$$

$$B^{(0)} = b \cosh(k_1 k_2 \tilde{t}) - 2a \frac{k_2}{k_1} \sinh(k_1 k_2 \tilde{t}), \tag{IV.39b}$$

where

$$k_1 = \left(\left(\frac{1}{2}\right) + \delta_1\right)^{1/2}, \qquad k_2 = \left(\left(\frac{1}{2}\right) - \delta_1\right)^{1/2}.$$

When $\delta_1 = -1/2$ we have

$$A^{(0)} = 2a, \qquad B^{(0)} = b - 2a\tilde{t}, \tag{IV.40a}$$

and for $\delta_1 = 1/2$ we obtain

$$A^{(0)} = 2a - b\tilde{t}, \qquad B^{(0)} = b. \tag{IV.40b}$$

Thus, the choice of $\tilde{t} = \epsilon t$ has led to the correct stability requirements (and this choice is unique). Furthermore, it can be verified that the expression we have derived for $F^{(0)}$ possesses the appropriate properties on, and in the neighborhoods of the two transitional curves through $\delta_c = 1/4$.

By carrying out the next step in the expansion we will obtain the more accurate criteria for $\delta^{(1)}$ and ${}^{(1)}\delta$ as well as the next term in the expansion for y.

The analysis for the third critical point, $\delta_c = 1$, is quite similar to the preceding with the exception that here we must set $\eta_\delta(\epsilon) = \epsilon^2$ and we use the notation $\tilde{\tilde{t}} = \epsilon^2 t$. We will not carry out the details here, but will simply present the final results for $F^{(0)}$.

We have for $\delta = 1 + \epsilon\delta_1 + \epsilon^2\delta_2 + O(\delta^3)$

$$y = \sum_{n=0}^{N} F^{(n)}(t, \tilde{\tilde{t}})\,\epsilon^n + O(\epsilon^{N+1}), \tag{IV.41}$$

$$F^{(0)} = A^{(0)}(\tilde{\tilde{t}})\sin t + B^{(0)}(\tilde{\tilde{t}})\cos t, \tag{IV.42a}$$

$$A^{(0)}(0) = a, \tag{IV.42b}$$

$$B^{(0)}(0) = b. \tag{IV.42c}$$

Boundedness with respect to t requires that $\delta_1 = 0$, and the conditions defining $A^{(0)}$ and $B^{(0)}$ are provided by the boundedness of $F^{(2)}$. The solutions are stable if $\delta_2 > 5/12$ or $\delta_2 < -1/12$ and

$$\begin{aligned} A^{(0)} = {} & \mp b\left(\frac{12\delta_2 - 5}{12\delta_2 + 1}\right)^{1/2} \sin\left[\left(\delta_2 - \frac{5}{12}\right)\left(\delta_2 + \frac{1}{12}\right)\right]^{1/2}\frac{\tilde{\tilde{t}}}{2} \\ & + a\cos\left[\left(\delta_2 - \frac{5}{12}\right)\left(\delta_2 + \frac{1}{12}\right)\right]^{1/2}\frac{\tilde{\tilde{t}}}{2}, \end{aligned} \tag{IV.43a}$$

$$\begin{aligned} B^{(0)} = {} & b\cos\left[\left(\delta_2 - \frac{5}{12}\right)\left(\delta_2 + \frac{1}{12}\right)\right]^{1/2}\frac{\tilde{\tilde{t}}}{2} \\ & \pm a\left(\frac{12\delta_2 + 1}{12\delta_2 - 5}\right)^{1/2} \sin\left[\left(\delta_2 - \frac{5}{12}\right)\left(\delta_2 + \frac{1}{12}\right)\right]^{1/2}\frac{\tilde{\tilde{t}}}{2}, \end{aligned} \tag{IV.43b}$$

where the upper or lower signs are to be taken depending on whether $\delta_2 > 5/12$ or $\delta_2 < -1/12$ respectively.

The solutions are unstable if $-1/12 < \delta_2 < 5/12$ and

$$A^{(0)} = b\frac{k_1}{k_2}\sinh k_1 k_2 \frac{\tilde{\tilde{t}}}{2} + a\cosh k_1 k_2 \frac{\tilde{\tilde{t}}}{2}, \tag{IV.44a}$$

$$B^{(0)} = b\cosh k_1 k_2 \frac{\tilde{\tilde{t}}}{2} + a\frac{k_2}{k_1}\sinh k_1 k_2 \frac{\tilde{\tilde{t}}}{2}, \tag{IV.44b}$$

where

$$k_1 = \left(\left(\frac{5}{12}\right) - \delta_2\right)^{1/2}, \qquad k_2 = \left(\left(\frac{1}{12}\right) + \delta_2\right)^{1/2}.$$

On the transitional curve $\delta^{(2)} = 1 + (5/12)\epsilon^2$, $A^{(0)}$ and $B^{(0)}$ take on the following values

$$A^{(0)} = a, \qquad B^{(0)} = \frac{a\tilde{\tilde{t}}}{4} + b. \tag{IV.45a}$$

On ${}^{(2)}\delta = 1 - (\epsilon^2/12)$ we have

$$A^{(0)} = \frac{b\tilde{\tilde{t}}}{4} + a, \qquad B^{(0)} = b. \tag{IV.45b}$$

Again, it can be verified that the expression we have derived for $F^{(0)}$ possesses the appropriate properties on, and in the neighborhood of the transitional curves through the critical point $\delta_c = 1$.

d. *Duffing's Equation, Modulated Harmonic Oscillations.* Consider the initial value problem

$$\ddot{y} + y + \epsilon y^3 = \epsilon\cos(1 + \alpha\epsilon)t, \tag{IV.46a}$$

$$y(0) = a, \tag{IV.46b}$$

$$\dot{y}(0) = b, \tag{IV.46c}$$

where α, a, and b are arbitrary constants. Note that by suitable normalization of y and t, any forcing function with an amplitude of order ϵ and a frequency differing from unity by order ϵ can be brought to the above form.

Since the variable $(1 + \alpha\epsilon)t$ appears directly in the forcing function of (IV.46a) we choose $t^+ = (1 + \alpha\epsilon)t$ as our fast variable and let $\tilde{t} = \epsilon t$ be the second variable.

Equations (IV.46) then transform to

$$(1+\alpha\epsilon)^2\frac{d^2y}{d(t^+)^2}+y+\epsilon y^3=\epsilon\cos t^+, \tag{IV.47a}$$

$$y(0)=a, \tag{IV.47b}$$

$$\left.\frac{dy}{dt^+}\right]_{t^+=0}=\frac{b}{(1+\alpha\epsilon)}. \tag{IV.47c}$$

We assume an expansion for y of the form

$$y=\sum_{n=0}^{N}F^{(n)}(t^+,\tilde{t})\,\epsilon^n+O(\epsilon^{N+1}) \tag{IV.48}$$

and derive the following equations for $F^{(0)}$ and $F^{(1)}$

$$F_{11}^{(0)}+F^{(0)}=0, \tag{IV.49a}$$

$$F_{11}^{(1)}+F^{(1)}=-2F_{12}^{(0)}-F^{(0)^3}+2\alpha F^{(0)}+\cos t^+. \tag{IV.49b}$$

With the solution for $F^{(0)}$ as follows

$$F^{(0)}=A(\tilde{t})\cos t^+ + B(\tilde{t})\sin t^+ \tag{IV.50}$$

where (IV.47) imply

$$A(0)=a, \qquad B(0)=b, \tag{IV.51}$$

(IV.49b) becomes

$$\begin{aligned} F_{11}^{(1)}+F^{(1)}=&\left[-2\frac{dB}{d\tilde{t}}-\frac{3}{4}A(A^2+B^2)+1+2\alpha A\right]\cos t^+ \\ &+\left[2\frac{dA}{d\tilde{t}}-\frac{3}{4}B(A^2+B^2)+2\alpha B\right]\sin t^+ \\ &+\frac{A}{4}(3B^2-A^2)\cos 3t^+ +\frac{B}{4}(B^2-3A^2)\sin 3t^+. \end{aligned} \tag{IV.52}$$

The requirement that $F^{(1)}$ be bounded gives the following two first order equations for A and B

$$\frac{dA}{d\tilde{t}}=\frac{3B}{8}(A^2+B^2)-\alpha B, \tag{IV.53a}$$

$$\frac{dB}{d\tilde{t}}=-\frac{3A}{8}(A^2+B^2)+\frac{1}{2}+\alpha A. \tag{IV.53b}$$

It is easy to verify that the system (IV.53) is Hamiltonian with

$$H = \frac{3}{32}(A^2 + B^2)^2 - \frac{A}{2} - \frac{\alpha}{2}(A^2 + B^2) = \text{const.} \tag{IV.54}$$

With this result A and B can be reduced to quadrature and are found to involve elliptic functions of $\tilde{t}$. The qualitative behavior of the integral curves in the (A, B) plane can best be deduced by considering (IV.53) directly.

We note that the singular points of this system are at $B = 0$ and A given by the roots of the cubic equation:

$$3A^3 - 8\alpha A - 4 = 0. \tag{IV.55}$$

If $\alpha < (81/128)^{1/3} \equiv \alpha_{CR}$, (IV.55) has only one positive root which we denote by A_0. If $\alpha > \alpha_{CR}$ we have, in addition to A_0 which remains positive, two negative roots $A_1 < A_2 < 0$.

The point $A = A_0$, $B = 0$ is a center for all values of α and, if $\alpha < \alpha_{CR}$, the integral curves form a nested set of closed simply connected curves which tend to large circles (cf. (IV.54)) for large values of H. The solution is sketched in Figure 1.

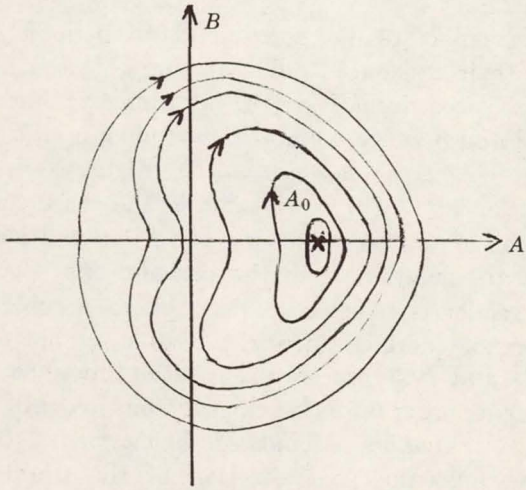


FIGURE 1. Amplitude-Phase Diagram for $\alpha < \alpha_{CR}$.

If $\alpha > \alpha_{CR}$, the point $A = A_1$, $B = 0$ is a saddle-point and $A = A_2$, $B = 0$ a center, and the integral curves have the qualitative behavior illustrated in Figure 2.

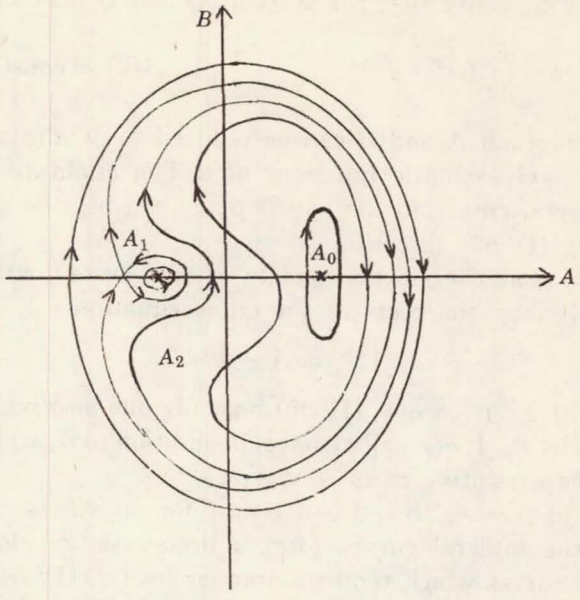


FIGURE 2. Amplitude-Phase Diagram for $\alpha > \alpha_{CR}$.

All three singular points correspond to periodic solutions of (IV.46) and their existence could, of course, have been deduced by Lindstedt's procedure. Here, in addition, we find the general behavior of A and B which we note is bounded in all cases. Furthermore, the sketch for the case $\alpha > \alpha_{CR}$ in Figure 2 shows that the periodic solution at $(A_1, 0)$ is unstable in the sense that any small deviation from the initial values $a = A_1$, $b = 0$ will lead to a large departure of the solution from the periodic one. For further discussion the reader is referred to [**25**], where Struble has studied this equation and derived identical results by his method.

Since $A(\tilde{t})$ and $B(\tilde{t})$ are transcendental functions, the calculation of the higher order terms becomes tedious and will not be given.

e. *Duffing's Equation, Modulated Subharmonic Oscillations*(9). Consider the following generalization of the problem discussed in §IVd:

$$\ddot{y} + y + \epsilon(y^3 + \mu\dot{y}) = \cos\Omega(\epsilon)\, t, \tag{IV.56a}$$

(9) The material in this section is taken from class notes prepared by P. A. Lagerstrom.

(IV.56b) $$y(0) = a,$$

(IV.56c) $$\dot{y}(0) = b,$$

where the frequency Ω is a known function of ϵ which may be expanded asymptotically in the form

(IV.57) $$\Omega(\epsilon) = \sum_{n=0}^{N} \omega_n \epsilon^n + O(\epsilon^{N+1}).$$

Thus, in addition to the forcing function of unit amplitude and arbitrary frequency we have included a linear damping term proportional to μ (where μ is a constant independent to ϵ).

There is no loss of generality in restricting ω_0 to nonnegative values since the solution for a given negative Ω is the same as that for $-\Omega$.

The case $\omega_0 = 0$ is a special case requiring separate treatment since here one should use Ωt as the slow variable. This case, which presents no special difficulties, will not be discussed.

The case $\omega_0 = 1$ also requires special consideration because the usual expansion for y would erroneously imply that $F^{(0)}$ is unbounded in response to a resonant forcing function (cf. (IV.58a) and (IV.59a)). The variable $\bar{y} = \epsilon^{-1/2} y$ and the perturbation problem in which $\bar{y}$ is regarded fixed as $\epsilon \to 0$ is the only meaningful one here (cf. §IVb where, for the linear problem, $\epsilon^{-1} y$ possessed a two variable expansion). However, since the nonlinear term now becomes $O(1)$ (as it should to prevent resonance) the problem is not within the scope of the present method.

In fact, we shall concentrate on the case $\omega_0 = 3$, which will be shown to lead to the fundamentally novel phenomenon of subharmonics.

We expand y in the form

(IV.58a) $$y = \sum_{n=0}^{N} F^{(n)}(t^+, \tilde{t})\, \epsilon^n + O(\epsilon^{N+1})$$

where

(IV.58b) $$t^+ = \Omega(\epsilon)\, t,$$

(IV.58c) $$\tilde{t} = \epsilon t^+$$

and compute the following equations for $F^{(0)}$ and $F^{(1)}$:

(IV.59a) $$\omega_0^2 F_{11}^{(0)} + F^{(0)} = \cos t^+,$$

(IV.59b) $$\omega_0^2 F_{11}^{(1)} + F^{(1)} = -2\omega_0\omega_1 F_{11}^{(0)} - 2\omega_0^2 F_{12}^{(0)} - \beta F_1^{(0)} - F^{(0)3}$$

where

$$\beta = \mu\omega_0.$$

With the notation

(IV.60a) $$\tau = \omega_0^{-1}(t^+ + \phi),$$

(IV.60b) $$\sigma = (1 - \omega_0^2)^{-1},$$

the solution of (IV.59a) may be written in the form

(IV.61) $$F^{(0)} = p\cos\tau + \sigma\cos t^+$$

where p and ϕ are as yet undefined functions of $\tilde{t}$.

The initial conditions imply that

(IV.62a) $$p(0) = [\omega_0^2 b^2 + (a - \sigma)^2]^{1/2} \equiv p_0,$$

(IV.62b) $$\phi(0) = \tan^{-1}\left(\frac{b\omega_0}{\sigma - a}\right) \equiv \phi_0.$$

After some manipulation (IV.59b) reduces to

(IV.63)
$$\begin{aligned}
\omega_0^2 F_{11}^{(1)} + F^{(1)} = &\left[2p\frac{d\phi}{d\tilde{t}} + 2p\frac{\omega_1}{\omega_0} - \frac{3}{2}p\sigma^2 - \frac{3}{4}p^3\right]\cos\tau \\
&+ \left[2\omega_0\frac{dp}{d\tilde{t}} + \frac{\beta}{\omega_0}p\right]\sin\tau \\
&- \frac{3}{4}p^2\sigma\cos(2\tau - t^+) \\
&+ \left(2\sigma\omega_0\omega_1 - \frac{3}{2}p^2\sigma - \frac{3}{4}\sigma^3\right)\cos t^+ \\
&+ \sigma\beta\sin t^+ - \frac{3}{4}p^2\sigma\cos(2\tau + t^+) \\
&- \frac{3}{4}p\sigma^2[\cos(2t^+ + \tau) + \cos(2t^+ - \tau)] \\
&- \frac{p^3}{4}\cos 3\tau - \frac{\sigma^3}{4}\cos 3t^+.
\end{aligned}$$

It is easy to verify that as long as $\omega_0 \neq 0, 1/3, 1$ or 3 only the terms with the argument τ (i.e., the first two bracketed terms on the right-hand side of (IV.63)) have the frequency ω_0^{-1} and hence lead to unbounded solutions for large values of t^+.

In this case p and ϕ are governed by

$$2\omega_0 \frac{dp}{d\tilde{t}} + \frac{\beta}{\omega_0} p = 0, \tag{IV.64a}$$

$$2p\frac{d\phi}{d\tilde{t}} + 2p\frac{\omega_1}{\omega_0} - \frac{3}{2} p\sigma^2 - \frac{3}{4} p^3 = 0, \tag{IV.64b}$$

whose solution is

$$p = p_0 e^{-(\beta/2\omega_0^2)\tilde{t}}, \tag{IV.65a}$$

$$\phi = \frac{3}{8}\frac{\omega_0^2 p_0^2}{\beta}(1 - e^{-(\beta/\omega_0^2)\tilde{t}}) + \left(\frac{3\sigma^2}{4} - \frac{\omega_1}{\omega_0}\right)\tilde{t} + \phi_0. \tag{IV.65b}$$

Thus, only the oscillations with integer multiples of the impressed frequency are unaffected by the damping (cf. (IV.61) and the right-hand side of (IV.63) with $p \to 0$). This common phenomenon of nonlinear oscillations is pointed out here to contrast the more unusual behavior to be exhibited later on for $\omega_0 = 3$.

If $\omega_0 = 1/3$, the term $-\sigma^3 \cos 3t^+/4$ has the same frequency ($\omega_0^{-1} = 3$) as the homogeneous solution of $F^{(1)}$ and must therefore be eliminated. It is easy to show that with $\omega_0 = 1/3$, $3t^+ = \tau - 3\phi$ hence the term in question will have equal contributions proportional to $\sin\tau$ and $\cos\tau$. When these contributions are added to the terms already considered in (IV.64) and ω_0 is set equal to $1/3$ one obtains

$$\frac{2}{3}\frac{dp}{d\tilde{t}} + 3\beta p - \frac{1}{4}\left(\frac{9}{8}\right)^3 \sin 3\phi = 0, \tag{IV.66a}$$

$$2p\frac{d\phi}{d\tilde{t}} + \left[6\omega_1 - \frac{3}{2}\left(\frac{9}{8}\right)^2\right] p - \frac{3}{4} p^3 - \frac{1}{4}\left(\frac{9}{8}\right)^3 \cos 3\phi = 0. \tag{IV.66b}$$

We note that for $\beta = 0$ the above system possesses the integral

$$H = \frac{3}{16} p^4 + 3\left[\frac{1}{4}\left(\frac{9}{8}\right)^2 - \omega_1\right] p^2 + \frac{1}{4}\left(\frac{9}{8}\right)^3 p \cos 3\phi \tag{IV.67}$$

which reduces the solution to quadrature.

Except for some trivial difference in numerical constants and the fact that in this case 3ϕ is to be regarded as the phase angle, the above system is identical to the one discussed in the previous section and will not be reconsidered here. The singular points of (IV.66), which define the periodic solutions to order unity, correspond to amplitudes at which steady-state oscillations with three times the impressed frequency are found.

We now consider the final distinguished value of $\omega_0 = 3$. It is easy to verify that this is the only other value of ω_0 for which a term in (IV.63) with an argument other than τ can resonate. In fact, if $\omega_0 = 3$, the term $-3p^2\sigma\cos(2\tau - t^+)/4$ becomes $3p^2\cos(\phi - \tau)/32$ and must be included in the grouping of terms with the frequency 1/3.

When the first two bracketed terms on the right-hand side of (IV.63) are augmented by the above contributions, the equations governing p and ϕ for the case $\omega_0 = 3$ become

$$\frac{dp}{d\tilde{t}} = -\frac{\beta p}{18} - \frac{p^2}{64}\sin\phi, \tag{IV.68a}$$

$$\frac{d\phi}{d\tilde{t}} = \frac{3}{8}p^2 + \left(\frac{3}{256} - \frac{\omega_1}{3}\right) - \frac{3p}{64}\cos\phi. \tag{IV.68b}$$

This is the most interesting value of ω_0 for this example, and will be considered in detail. We note here that if, unlike (IV.65a), one finds that $p \to p_\infty \neq 0$ as $\tilde{t} \to \infty$ with $\beta > 0$, then (cf. (IV.60a) and (IV.61)) the phenomenon of "subharmonic" oscillations is demonstrated. For, in this event the forcing function will have excited one-third its frequency in the term $p\cos\tau$ of (IV.61).

The notation

$$r = 8p, \tag{IV.69a}$$

$$\gamma = \frac{256}{9}\beta, \tag{IV.69b}$$

$$\omega_1 = \frac{9}{512}\left(\frac{7}{4} + \nu\right), \tag{IV.69c}$$

$$T = \frac{\tilde{t}}{512}, \tag{IV.69d}$$

simplifies (IV.68) to

$$\frac{dr}{dT} = -r(r\sin\phi + \gamma), \tag{IV.70a}$$

$$\frac{d\phi}{dT} = 3\left(r^2 - r\cos\phi + \frac{1}{4} - \nu\right). \tag{IV.70b}$$

It is easily verified that if $\gamma = 0$, i.e., zero damping, the system (IV.70) possesses the integral

$$H = r^3\cos\phi - \frac{3}{4}r^4 + \frac{3}{2}r^2\left(\nu - \frac{1}{4}\right) \tag{IV.71}$$

which shows that $F^{(0)}$ describes bounded oscillations. Moreover, since the only damping contribution is in the term $-\gamma r$ of (IV.66a), we conclude that the effect of $\gamma > 0$ is to decrease a given value of r locally. Although the solution of (IV.70) can be brought to quadrature if $\gamma = 0$, it is more instructive, even for this case, to study the qualitative nature of the integral curves. For this purpose we set

$$\xi = r\cos\phi, \tag{IV.72a}$$

$$\eta = r\sin\phi, \tag{IV.72b}$$

and transform (IV.70) to

$$\frac{d\xi}{dT} = -\xi(\eta + \gamma) - 3\eta\left[\left(\xi - \frac{1}{2}\right)^2 + \eta^2 - \nu\right], \tag{IV.73a}$$

$$\frac{d\eta}{dT} = -\eta(\eta + \gamma) + 3\xi\left[\left(\xi - \frac{1}{2}\right)^2 + \eta^2 - \nu\right]. \tag{IV.73b}$$

The singular points of (IV.73) which we denote by (ξ_i, η_i), $i = 1, 2, 3$ are easily computed to be

$$\xi_1 = \eta_1 = 0, \tag{IV.74a}$$

$$\eta_2 = \eta_3 = -\gamma, \tag{IV.74b}$$

$$\xi_2 = \frac{1}{2} + (\nu - \gamma^2)^{1/2}, \qquad \xi_3 = \frac{1}{2} - (\nu - \gamma^2)^{1/2}. \tag{IV.74c}$$

We note from (IV.74c) that the singular points away from the origin can only exist if $\nu \geqq \gamma^2$, which according to (IV.69c) means that

$$\omega_1 \geqq \frac{9}{512}\left(\frac{7}{4}+\gamma^2\right) \equiv \omega_{CR}.$$

The singular point at the origin, which always exists, is a spiral (or center if $\gamma = 0$) and linearization of (IV.73) in the neighborhood of the origin leads to the following approximate solution:

(IV.75a) $$r \doteq r_0 e^{-\gamma T},$$

(IV.75b) $$\phi \doteq 3\left(\frac{1}{4}-\nu\right) T + \phi_0,$$

where r_0 and ϕ_0 are constants of integration.

Thus, if $\omega < \omega_{CR}$ the behavior described by (IV.75) persists, qualitatively at least, for all values of r. If $\gamma > 0$, the integral curves are spirals which tend towards the origin, while for $\gamma = 0$ they form a set of nested, closed, simply-connected curves which, for large distances, tend to circles centered at the origin.

To study the more interesting case $\omega_1 > \omega_{CR}$ we set $\nu = a^2$ and denote the ξ, η coordinates, referring to the two additional singular points defined in (IV.74b) and (IV.74c), by an asterisk, i.e.,

(IV.76a) $$\xi_i^* = \xi - \xi_i,$$

(IV.76b) $$\eta_i^* = \eta - \eta_i.$$

Equations (IV.69), when linearized in the neighborhoods of the above singular points become

(IV.77a) $$\frac{d\xi_i^*}{dT} = -\xi_i\eta_i^* + 6\gamma\left[\left(\xi_i - \frac{1}{2}\right)\xi_i^* - \gamma\eta_i^*\right],$$

(IV.77b) $$\frac{d\eta_i^*}{dT} = \gamma\eta_i^* + 6\xi_i\left[\left(\xi_i - \frac{1}{2}\right)\xi_i^* - \gamma\eta_i^*\right].$$

For $\gamma = 0$, we distinguish the two cases $a \gtrless 1/2$. If $a > 1/2$, $\xi_3 = (1/2) - a < 0$ while if $a < 1/2$, both ξ_2 and ξ_3 are on the positive ξ axis. In either case $(\xi_2, 0)$ is a center while $(\xi_3, 0)$ is a saddle and the integral curves have the qualitative behavior sketched in Figures 3 and 4.

If $\gamma > 0$, the two centers become spirals, while the saddle-point persists, and one can deduce the qualitative behavior of the integral curves shown in Figures 5 and 6.

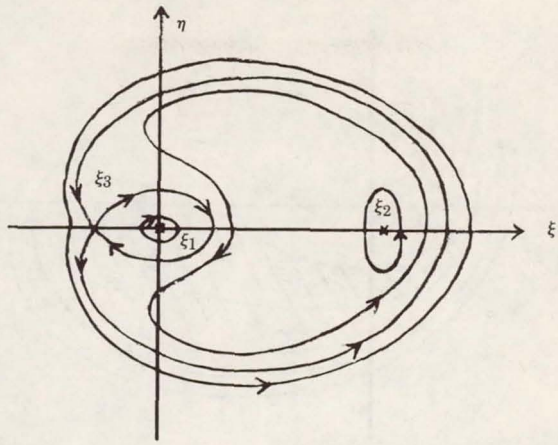


FIGURE 3. Amplitude-Phase Diagram for $a > \frac{1}{2}$, $\gamma = 0$.

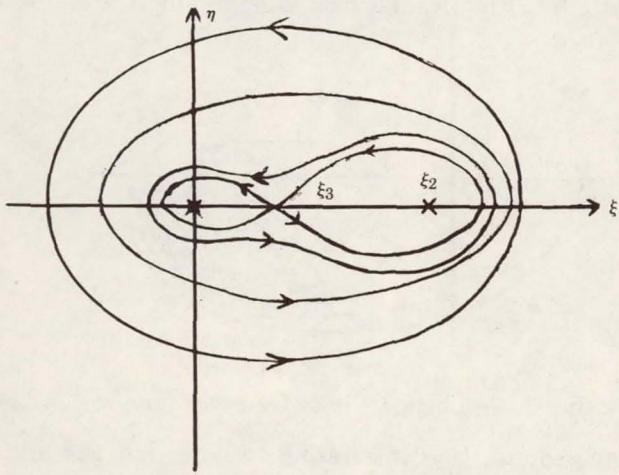


FIGURE 4. Amplitude-Phase Diagram for $a < \frac{1}{2}$, $\gamma = 0$.

Figures 5 and 6 confirm the fact that for certain initial amplitudes the solution tends to the steady-state subharmonic oscillations with the amplitude $\xi_2 = (1/2) + (\nu - \gamma^2)^{1/2}$.

The shaded regions in Figures 5 and 6 correspond to solutions which eventually achieve subharmonic oscillation while the unshaded regions correspond to solutions whose subharmonic compo-

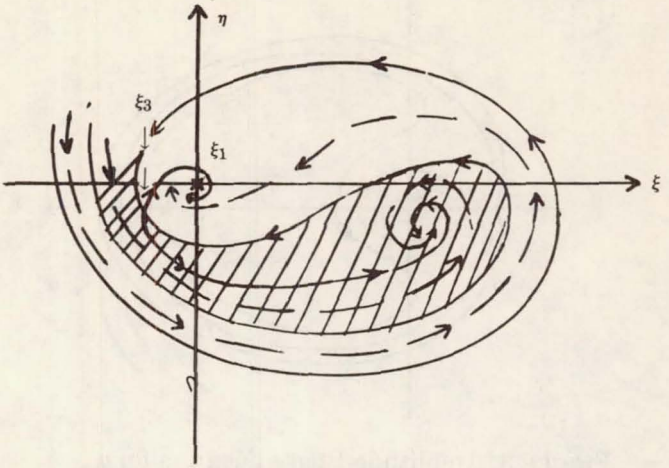


FIGURE 5. Amplitude-Phase Diagram for $a > \frac{1}{2}, \gamma > 0$.

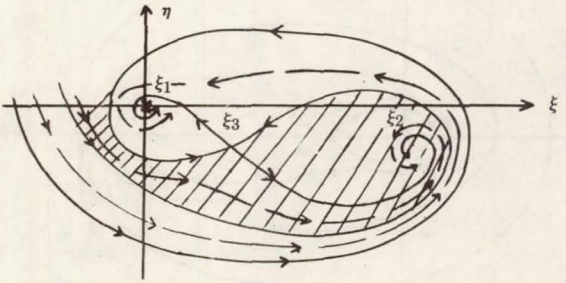


FIGURE 6. Amplitude-Phase Diagram for $a < \frac{1}{2}, \gamma > 0$.

nent is damped out. It is interesting to note that the initial values which lead to one or the other of the above stable modes are located on spiralling bands whose actual values can only be determined by detailed numerical computation.

f. *Multiple Damping.* In order to emphasize the crucial role that the choice of variables plays in the success or failure of the two variable expansion procedure, we will briefly consider the following equation

$$\ddot{y} + y + \epsilon[\dot{y}^3 + 3\epsilon\nu\dot{y}] = 0 \tag{IV.78}$$

where ν is a positive constant independent of ϵ.

This equation was pointed out by Morrison in [**20**] as an example where the direct application of the two variable method fails, and indeed this is easily verified by routinely expanding the solution in the form (II.23) used in §II.

Since the only difference between the perturbing function in (IV.78) and that discussed in §IIIb is the term $3\epsilon\nu\dot{y}$, one will duplicate the results found for the problem of §IIIb to $O(1)$. Furthermore, an easy calculation shows that the added term in (IV.78) only modifies the equation for $B^{(1)}$ to the following extent (cf. (III.10b))

$$2\frac{dB^{(1)}}{d\tilde{t}} + \frac{9}{4}B^{(0)2}B^{(1)} + 3\nu B^{(0)} = 0. \tag{IV.79}$$

Using the result given in (III.6b) for $B^{(0)}$, (IV.79) can be readily integrated to yield

$$B^{(1)} = \frac{1}{(3\tilde{t}+4)^{3/2}}\left[8B^{(1)}(0) - \frac{9}{2}\nu\tilde{t}^2 - 12\nu\tilde{t}\right]. \tag{IV.80}$$

The inconsistency of the two variable expansion is depicted by the terms proportional to $\tilde{t}$ and $\tilde{t}^2$ in the brackets of (IV.80). No formal means of eliminating these terms is available within the framework of the assumed expansion. Incidentally, the term $(9/2)\nu\tilde{t}^2(3\tilde{t}+4)^{-3/2}$ is not only inconsistent, but also $O(\tilde{t}^{1/2})$ as $\tilde{t}\to\infty$.

As pointed out by Morrison, the implicit assumption in the present expansion that $3\epsilon^2\nu\dot{y} = O(\epsilon^2)$ while $\epsilon\dot{y}^3 = O(\epsilon)$ is not true for sufficiently small amplitudes, i.e., for sufficiently long times. Herein also lies the answer to the correct choice of variables for the applicability of the two variable expansion.

Since in the domain in t which is of interest $\epsilon\dot{y}^3 = O(\epsilon^2\dot{y})$, we must use a "blown-up" variable $\bar{y} = \epsilon^{-1/2}y$ in terms of which the two terms in question are explicitly of the same order of magnitude.

Thus, (IV.78) becomes

$$\frac{d^2\bar{y}}{dt^2} + \bar{y} + \epsilon^2\left[\left(\frac{d\bar{y}}{dt}\right)^3 + 3\nu\frac{d\bar{y}}{dt}\right] = 0 \tag{IV.78′}$$

and the method discussed in §II now applies directly as long as ϵ^2 is used instead of ϵ in all the formulas.

In particular, the correct slow variable for this problem is $\epsilon^2 t$ and a consistent, bounded solution for $\overline{y}$ in terms of the two variables t and $\epsilon^2 t$ can be derived routinely using the formulas of §II. One can also verify that as $\nu \to 0$, the results so derived reduce to those of §IIIb.

g. *Nonuniformity with Respect to an Initial Parameter.* In the preceeding example it was shown how the two variable expansion, in terms of a given set of variables, might break down because the orders of magnitude assigned by the assumed expansion to two of the perturbation terms is incorrect for times which are sufficiently large.

In this section we will study an example where the orders of magnitude of two perturbation terms is in error not for long times, but for certain values of an initial parameter say the amplitude. This more subtle breakdown of the two variable expansion procedure and the correct treatment of the problem for the case of a critical initial amplitude will be only summarized here. For a detailed exposition as well as a discussion of the related problem in celestial mechanics, the reader is referred to [6].

Consider the equation

$$\ddot{y} + y + 2\epsilon y(1 - 5\cos^2 R) = \epsilon^2 R \cos t \tag{IV.81a}$$

where

$$R^2 = y^2 + \dot{y}^2. \tag{IV.81b}$$

In the absence of the forcing function, this equation can be integrated exactly and exhibits the following behavior in the phase-plane of y and $\dot{y}$. Whenever the radius $R = [y^2 + \dot{y}^2]^{1/2}$ in the phase-plane takes on the critical values $R_c = \cos^{-1}(5)^{-1/2}$, the motion reduces to simple harmonic oscillations with amplitude R_c and unit frequency. For each annular region bounded by two consecutive values of R_c, the integral curves are ovals with their axes aligned alternately parallel either to y or to $\dot{y}$. One would thus expect that the addition of the forcing term with unit frequency will cause local resonance in neighborhoods of the critical amplitudes R_c. As will be shown later on in this section, this will indeed be the case and will give rise to the problem of the "critical amplitude."

We assume that y may be expanded in terms of t and $\tilde{t} = \epsilon t$

in the form([10])

$$y = \sum_{n=0}^{N} F^{(n)}(t, \tilde{t}; \epsilon)\, \epsilon^n + O(\epsilon^{N+1}). \tag{IV.82}$$

The governing equation for $F^{(0)}$,

$$F^{(0)}_{11} + F^{(0)} = 0 \tag{IV.83}$$

has the solution

$$F^{(0)} = \alpha(\tilde{t}; \epsilon) \cos[t - \beta(\tilde{t}; \epsilon)] \tag{IV.84}$$

where we assume that α and β may also be expanded asymptotically in the form

$$\alpha(\tilde{t}; \epsilon) = \sum_{n=0}^{N} \alpha_n(\tilde{t})\, \epsilon^n + O(\epsilon^{N+1}), \tag{IV.85a}$$

$$\beta(\tilde{t}; \epsilon) = \sum_{n=0}^{N} \beta_n(\tilde{t})\, \epsilon^n + O(\epsilon^{N+1}). \tag{IV.85b}$$

From (IV.81) and the above, the following equation for $F^{(1)}$ can be calculated

$$\begin{aligned} F^{(1)}_{11} + F^{(1)} &= 2\frac{d\alpha_0}{d\tilde{t}} \sin(t - \beta) \\ &\quad - 2\alpha_0 \left[\frac{d\beta_0}{d\tilde{t}} + (1 - 5\cos^2\alpha_0) \right] \cos(t - \beta). \end{aligned} \tag{IV.86}$$

The boundedness of $F^{(1)}$ gives

$$\alpha_0 = a_0, \tag{IV.87a}$$

$$\beta_0 = s_0\tilde{t} + b_0, \tag{IV.87b}$$

where a_0 and b_0 are constants depending upon the initial conditions and

$$s_0 = 5\cos^2\alpha_0 - 1. \tag{IV.88}$$

With the right-hand side of (IV.86) equal to zero, the solution

([10]) In (IV.82) the fact that the $F^{(n)}$ may also depend on ϵ is a trivial generalization of the previous procedure to allow for routine computational simplifications.

$F^{(1)} = 0$ introduces no loss of generality by virtue of the expansions (IV.85).

The equation for $F^{(2)}$ can be calculated in the form

$$F_{11}^{(2)} + F^{(2)} = \left(2\frac{d\alpha_1}{d\tilde{t}} - \alpha_0 \sin\beta\right) \sin(t-\beta)$$

$$\text{(IV.89a)} \qquad + \alpha_0 \left(s_0^2 - \frac{5}{2}\alpha_0 s_0 \sin 2\alpha_0 + \cos\beta + 2s_0'\alpha_1 - 2\frac{d\beta_1}{d\tilde{t}}\right) \cos(t-\beta) - \frac{5}{2}\alpha_0^2 s_0 (\sin 2\alpha_0) \cos 3(t-\beta)$$

where

$$\text{(IV.89b)} \qquad s_0' \equiv \frac{ds_0}{d\alpha_0} = -5\sin 2\alpha_0.$$

The requirement that $F^{(2)}$ be bounded dictates the following equations for α_1 and β_1:

$$\text{(IV.90a)} \qquad \frac{d\alpha_1}{d\tilde{t}} = \frac{\alpha_0}{2}\sin\beta,$$

$$\text{(IV.90b)} \qquad \frac{d\beta_1}{d\tilde{t}} = \frac{s_0^2}{2} - \frac{5}{4}\alpha_0 s_0 \sin 2\alpha_0 + \frac{1}{2}\cos\beta + s_0'\alpha_1.$$

Since for $s_0 \to 0$ (i.e., $\alpha_0 = \cos^{-1}(5)^{-1/2} \equiv \alpha_c$) $\beta = b_0 + \epsilon\beta_1 + O(\epsilon^2)$, the integration of (IV.90) will lead to the conclusion that $\alpha_1 = O(\epsilon^{-1})$ for this case. This is an inconsistency([11]) in the expansion procedure indicative of a nonuniformity as $s_0 \to 0$. We point out, however, that as long as s_0 is not small the integration of (IV.90) leads to consistent and well-behaved results.

In this simple model the cause of the difficulty is easy to discern and remedy. As was pointed out earlier, whenever $\alpha = \alpha_c$ the nonlinear system degenerates to simple harmonic motion at the frequency of the forcing function. Therefore, in some neighborhood of α_c the amplitude must locally behave as though the system was in linear resonance, (cf. the discussion after (IV.90b)). Due to this local effect the forcing function, which would otherwise be of order ϵ^2 now takes on a more important role. This fact is exhibited in

([11]) Alternatively, if in (IV.90a) we regard $\sin\beta = \sin b_0 + O(\epsilon)$, we conclude $\alpha_1 = \tilde{t}\,(\alpha_0/2)\sin b_0$ which is also inconsistent.

the equations for $d\alpha_0/d\tilde{t}$ and $d\beta_0/d\tilde{t}$ which are both small when s_0 is small. Hence in this case one cannot neglect the higher order terms in computing α_0 and β_0.

This observation was first made by Struble in [26] in connection with the related problem of a "critical" inclination in celestial mechanics. The intuitively obvious approach for the case $|s_0| \ll 1$ is therefore the anticipation of the importance of the forcing function and its immediate involvement in the solution for α_0 and β_0. Thus, we append $\epsilon\alpha\cos t = \epsilon\alpha\,[\cos(t-\beta)\cos\beta - \sin(t-\beta)\sin\beta\,]$ which is the contribution to $O(\epsilon)$ of the right-hand side of (IV.81a) to (IV.86) and obtain the following equations([12]) for α and β:

$$\frac{d\alpha}{d\tilde{t}} = \frac{\epsilon}{2}\,\alpha\sin\beta, \tag{IV.91a}$$

$$\frac{d\beta}{d\tilde{t}} = -\,(1 - 5\cos^2\alpha) + \frac{\epsilon}{2}\cos\beta. \tag{IV.91b}$$

The terms of order ϵ in (IV.91) which are exclusively the contributions of the right-hand side of (IV.81a) will radically alter the behavior of α and β near the critical amplitudes.

Equations (IV.91) are Hamiltonian, hence along an integral curve

$$2H = 3\alpha + \frac{5}{2}\sin 2\alpha + \epsilon\alpha\cos\beta = \text{const.} \tag{IV.92}$$

With the aid of (IV.92), the integral curves in the α, β plane can be easily calculated. The singular points are located at $\beta = \beta_s = n\pi$, $n = 0, \pm 1, \pm 2, \cdots$ and $\cos\alpha = \cos\alpha_s = \pm\,((2 \mp \epsilon)/10)^{1/2}$.([13]) These points form an alternating pattern of centers and saddle-points with solution curves as shown qualitatively in Figure 7. It should be noted that for each of the infinite number of values of α with $\cos\alpha$ near to $\pm\ 5^{-1/2}$ there is a stream of solutions analogous to the stream which is shown in Figure 7 for α in the first quadrant and $\cos\alpha$ near $5^{-1/2}$.

([12]) Note the subscripts on α and β have been dropped in the ensuing discussion to distinguish this case from the previous calculations where $s_0 \neq 0$. However, α and β, as used here, should be regarded as approximations of the quantities appearing in (IV.84) for the case $|s_0| \ll 1$.

([13]) The upper or lower signs inside the radical are to be taken when β is an even or odd multiple of π respectively.

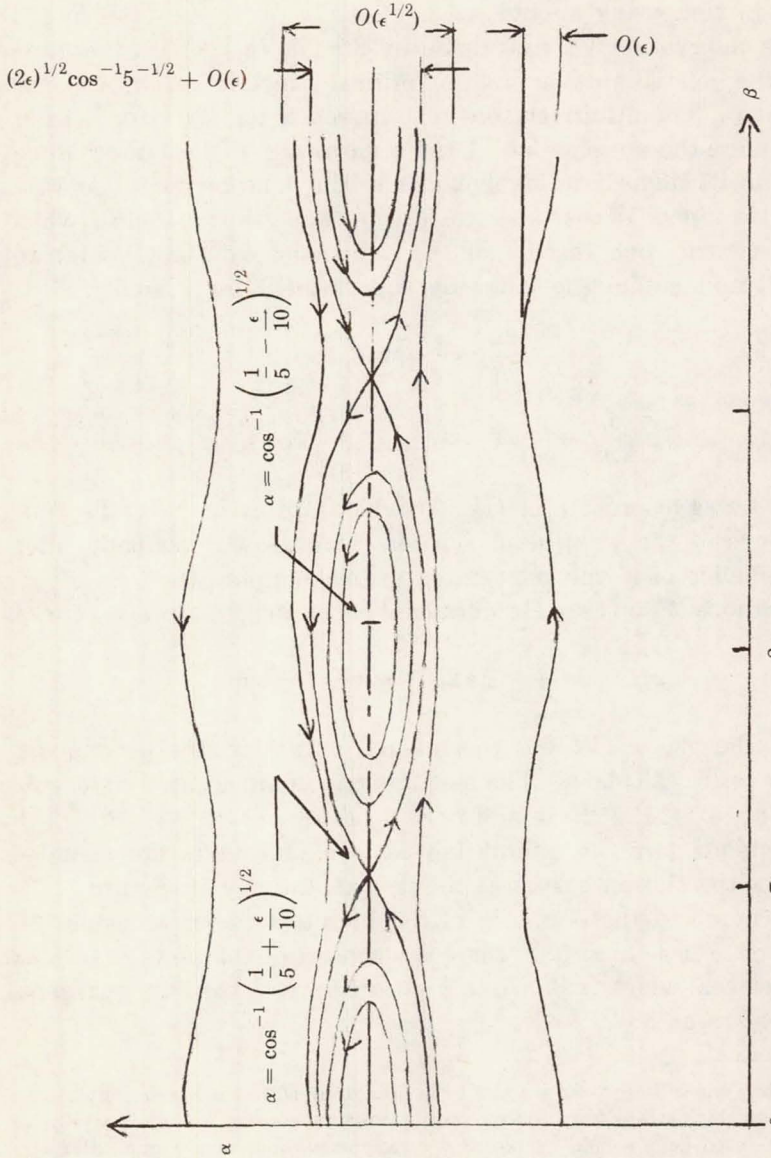


FIGURE 7. Amplitude-Phase Diagram Showing a Critical Amplitude.

We observe three possible types of motion if we consider the integral curves in horizontal strips with a width of order $\epsilon^{1/2}$ centered about any of the critical amplitudes.

The integral curves which pass through two adjacent saddle-points for a given value of α_s form the boundaries of oval regions with a width also of $O(\epsilon^{1/2})$ inside which both α and β undergo bounded oscillations. For example, the motion in the neighborhood of the point $\beta_s = 0$ and $\alpha_s = \cos^{-1}((2-\epsilon)/10)^{1/2}$ has the form

$$\alpha = \alpha_s + c_1\epsilon^{1/2}\cos\left[(2\epsilon\alpha_s)^{1/2}\tilde{t} + c_2\right], \tag{IV.93a}$$

$$\beta = -4c_1(2\alpha_s)^{-1/2}\sin\left[(2\epsilon\alpha_s)^{1/2}\tilde{t} + c_2\right], \tag{IV.93b}$$

where c_1 and c_2 are small constants depending on initial conditions which allow the linearization of (IV.91).

The separatrix forming the above boundary corresponds to motion where α and β approach the value at the saddle-point asymptotically as $\tilde{t} \to \infty$. In fact, by use of (IV.92) it is easy to show that the separatrix around the point $\beta_s = 0$ and $\alpha_s = \cos^{-1}((2-\epsilon)/10)^{1/2}$ for $0 < \alpha_s < \pi/2$ intersects the α axis at a distance $(\epsilon/2)^{1/2}\cos^{-1}(5)^{-1/2} + O(\epsilon)$ from the singular point. Finally, the motion just outside the oscillatory regions is characterized by the fact that α undergoes bounded oscillations, while β has a secular motion superimposed on its oscillations. In all three of the above motions the characteristic frequency is $O(\epsilon^{3/2})$ in the natural time variable whereas the amplitudes of oscillation are $O(\epsilon^{1/2})$ (cf. (IV.93)). This immediately suggests that the slow time scale appropriate for motion near the critical amplitudes is $\tilde{t} = \epsilon^{3/2}t$, and that one must seek an expansion for y in powers of $\epsilon^{1/2}$.

In contrast, when the initial amplitude is not critical, we note from (IV.90a) that α oscillates with amplitude and frequency both of order ϵ and that the oscillatory as well as secular components of β behave similarly.

The above intuitive construction essentially duplicates the results of the more systematic approach wherein one sets $s_0 = \epsilon^{1/2}\bar{s}_0$ holding $\bar{s}_0$ fixed as $\epsilon \to 0$ and uses $\tilde{t} = \epsilon^{3/2}t$ as the slow variable. The details of this as well as the important question of matching the two types of solutions near and away from the critical amplitudes are covered in [6].

V. **Applications for satellite motion.**

a. *Formulation of the Satellite Problem.* We consider the motion of a particle under the influence of an arbitrary force **F**. We assume that the variables are all made dimensionless by dividing distances by a given characteristic length, and the time by a given characteristic value to be specified for each problem. If x, y, z is a Cartesian inertial coordinate system we introduce the conventional spherical polar coordinates r, ψ, θ defined by

$$x = r\cos\psi\sin\theta, \tag{V.1a}$$

$$y = r\sin\psi\sin\theta, \tag{V.1b}$$

$$z = r\cos\theta. \tag{V.1c}$$

Let the vector from the origin to the particle be denoted by **x**. The components of generalized force in the r, ψ, θ directions are defined by

$$Q_r = \mathbf{F}\cdot\frac{\partial\mathbf{x}}{\partial r}, \tag{V.2a}$$

$$Q_\psi = \mathbf{F}\cdot\frac{\partial\mathbf{x}}{\partial\psi}, \tag{V.2b}$$

$$Q_\theta = \mathbf{F}\cdot\frac{\partial\mathbf{x}}{\partial\theta}. \tag{V.2c}$$

With the kinetic energy T in the form

$$T = \frac{1}{2}\left[\left(\frac{dr}{dt}\right)^2 + r^2\sin^2\theta\left(\frac{d\psi}{dt}\right)^2 + r^2\left(\frac{d\theta}{dt}\right)^2\right] \tag{V.3}$$

the equations of motion[14] may be written in the form

$$\frac{d}{dt}\left(r^2\sin^2\theta\frac{d\psi}{dt}\right) = Q_\psi, \tag{V.4a}$$

$$\frac{d}{dt}\left(r^2\frac{d\theta}{dt}\right) - r^2\sin\theta\cos\theta\left(\frac{d\psi}{dt}\right)^2 = Q_\theta, \tag{V.4b}$$

$$\frac{d^2r}{dt^2} - r\left(\frac{d\theta}{dt}\right)^2 - r\sin^2\theta\left(\frac{d\psi}{dt}\right)^2 = Q_r. \tag{V.4c}$$

---

[14] If x, y, z is not an inertial frame, we may still use (V.4) as long as we include in **F** and the ensuing values of Q_r, Q_ψ and Q_θ, the necessary fictitious forces due to the motion of the origin relative to the inertial center.

Since the satellite can be considered to move in an instantaneous plane defined by the distance and velocity vectors, one may also define the motion in terms of the following variables proposed by Struble in [24]. The geometry is sketched in Figure 8.

i = inclination; angle between x, y plane and the instaneous orbital plane.

Ω = longitude of the node; angle in the x, y plane between some fixed direction, say x and the ascending node.

ϕ = angle between the ascending node and the distance vector.

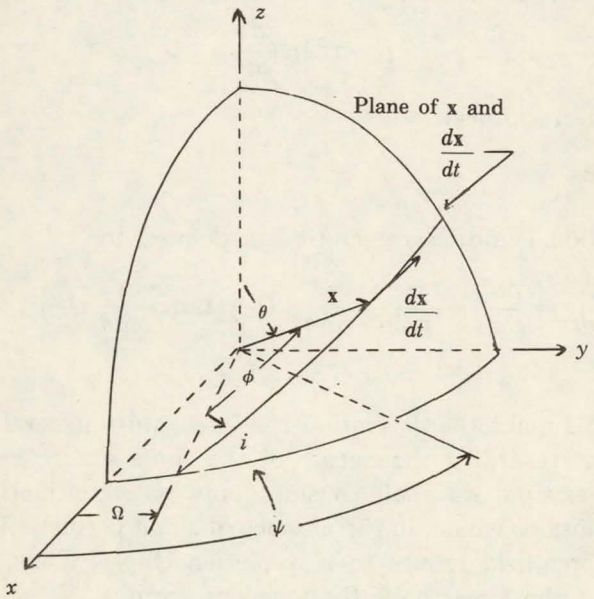


FIGURE 8. Geometry of Struble's Variables.

Struble([15]) has shown that equations (V.4) transform to the following fifth-order system after elimination of the time (where we use a prime to denote differentiation with respect to ϕ):

$$\text{(V.5a)} \qquad p' = Q_\psi \left[\frac{pu^2}{\cos i} + \frac{\cos^3 i \cos\theta}{p \sin^2 i \sin\theta} \left(Q_\theta + \tan i \frac{\cos\phi}{\sin\theta} Q_\psi \right) \right]^{-1},$$

$$\text{(V.5b)} \qquad \Omega' = - \frac{(Q_\theta + \tan i\,(\cos\phi/\sin\theta)\, Q_\psi)\, \cos^3 i \cos\theta}{p^2 u^2 \sin^2 i \sin\theta + \cos^4 i \cos\theta\,(Q_\theta + \tan i\,(\cos\phi/\sin\theta)\, Q_\psi)},$$

([15]) Note that Struble in [24] defines the node in the opposite sense.

$$\text{(V.5c)} \qquad i' = -\frac{\sin^2 i \cos^3 i \cos\phi\,(Q_\theta + \tan i\,(\cos\phi/\sin\theta)\,Q_\psi)}{p^2u^2\sin^2 i \sin\theta + \cos^4 i \cos\theta\,(Q_\theta + \tan i\,(\cos\phi/\sin\theta)\,Q_\psi)},$$

$$\text{(V.5d)} \qquad u'' - \frac{2}{u}(u')^2 + \frac{u'(d\phi/dt)'}{d\phi/dt} + \frac{(p^2u^5/\cos^2 i) + u^2 Q_r}{(d\phi/dt)^2} = 0,$$

where p is the component of angular momentum along the z axis and is defined by

$$\text{(V.5e)} \qquad p = r^2\sin^2\theta\frac{d\psi}{dt}$$

and u is the reciprocal radius

$$\text{(V.5f)} \qquad u = \frac{1}{r}.$$

In equation (V.5d) $d\phi/dt$ and θ are defined by

$$\text{(V.6a)} \qquad \frac{d\phi}{dt} = \frac{pu^2}{\cos i} + \frac{\cos^3 i \cos\theta}{p\sin^2 i \sin\theta}\left(Q_\theta + \tan i\frac{\cos\phi}{\sin\theta}\,Q_\psi\right),$$

$$\text{(V.6b)} \qquad \cos\theta = \sin i \sin\phi.$$

Up to this point the derivation has been quite general with no assumptions regarding the nature of the force **F**.

In this section we shall consider only satellite motions, i.e., bounded motions which, in the absence of small perturbations upon a central force field reduce to a Keplerian ellipse. Thus, we shall require that the forces have the following form

$$\text{(V.7a)} \qquad Q_r = -\frac{1}{r^2} + \epsilon\overline{Q_r},$$

$$\text{(V.7b)} \qquad Q_\psi = \epsilon\overline{Q_\psi},$$

$$\text{(V.7c)} \qquad Q_\theta = \epsilon\overline{Q_\theta},$$

where the $\overline{Q}$ are arbitrary functions([16]) of the coordinate velocities, t and ϵ restricted only to the extent that for a sufficiently small initial energy the particle will remain in some neighborhood of the

([16]) Note that by a proper choice of dimensionless variables, the dominant component of the gravitational attraction which is spherically symmetric can always be brought to the form $-1/r^2$, free of superfluous constants.

origin under the influence of the above forces.

This excludes, for example, a constant tangential thrust force, no matter how small its magnitude might be, since eventually such a force would impart a sufficient energy to propel the particle to escape velocity.

Use of (V.7) in (V.5a), (V.5b) and (V.5c) displays the fact that p, Ω and i are constants when $\epsilon = 0$. The important role of the above coordinate system becomes evident only after one substitutes (V.7) into (V.6b) and (V.5d) to obtain the following equation for u

$$u'' + (1+\epsilon H)^{-2}u = \left(1 - \epsilon\frac{\overline{Q_r}}{u^2}\right)\left(1+\epsilon H\right)^{-1}\frac{\cos^2 i}{p^2} + K \tag{V.8a}$$

where

$$H = \frac{M\cos i}{pu^2}, \tag{V.8b}$$

$$M = \frac{\cos^3 i\cos\theta}{p\sin^2 i\sin\theta}\left(\overline{Q_\theta} + \tan i\frac{\cos\phi}{\sin\theta}\,\overline{Q_\psi}\right), \tag{V.8c}$$

$$K = -\frac{uu'(u^2p'\cos i + pu^2i'\sin i + \epsilon M'\cos i) - 2\epsilon(u')^2M\cos i}{u\cos i(pu^2 + \epsilon M\cos i)} = O(\epsilon). \tag{V.8d}$$

Thus, as $\epsilon \to 0$, u is governed by the equation

$$u'' + u = \text{constant} \tag{V.9}$$

and the two variable expansion procedure applies to the problem with $\epsilon \ll 1$.

For planar motion, i.e., $\overline{Q_\theta} = \Omega = i = 0$; $\theta = \pi/2$; $\psi = \phi$; $\overline{Q_\psi} = \overline{Q_\phi}$, one obtains the following very simple equations

$$u'' + u = \frac{1}{p^2} - \frac{\epsilon}{p^2u^2}(\overline{Q_r} + \overline{Q_\phi}u'), \tag{V.10a}$$

$$p' = \epsilon\frac{\overline{Q_\phi}}{pu^2}, \tag{V.10b}$$

$$t' = \frac{1}{pu^2}. \tag{V.10c}$$

Equations (V.10) may also be written in the form given below where the value of p given by (V.10c) has been used explicitly

$$u'' + u = u^4(t')^2 - \epsilon u^2(t')^2 \left(\overline{Q_r} + \overline{Q_\phi} \frac{du}{d\phi} \right), \tag{V.11a}$$

$$(u^2 t')' = -\epsilon u^4 (t')^3 \overline{Q_\phi}. \tag{V.11b}$$

In the next two sections we shall consider two examples of planar motions. For applications of the two variable procedure to general nonplanar satellite orbits the reader is referred to [6] and [7].

b. *Aerodynamic Perturbations of a Satellite Orbit*([17]). A satellite is assumed to move in a planar orbit about a homogeneous, spherical, and nonrotating earth surrounded by a thin constant density atmosphere. It is also assumed that no other celestial bodies influence the motion of the satellite. The drag coefficient and lift-drag ratio of the satellite are taken to be constants. This physical model admittedly departs considerably from realism, and is presented only to illustrate the qualitative behavior of the motion. The more realistic model with a variable density, and aerodynamic coefficients depending on Mach number and angle of attack can be treated by this method at the cost of considerable algebraic complication.

The dimensional equations of motion may be written as

$$\frac{d^2 r^*}{dt^{*2}} - r^* \left(\frac{d\phi}{dt^*} \right)^2 = -\frac{GM}{r^{*2}} - \frac{C_D \rho S}{2m} \left[\frac{dr^*}{dt^*} - n r^* \frac{d\phi}{dt^*} \right] \left[\left(\frac{dr^*}{dt^*} \right)^2 + r^{*2} \left(\frac{d\phi}{dt^*} \right)^2 \right]^{1/2}, \tag{V.12a}$$

$$r^* \frac{d^2\phi}{dt^{*2}} + 2 \frac{dr^*}{dt^*} \frac{d\phi}{dt^*} = -\frac{C_D \rho S}{2m} \left[r^* \frac{d\phi}{dt^*} + n \frac{dr^*}{dt^*} \right] \left[\left(\frac{dr^*}{dt^*} \right)^2 + r^{*2} \left(\frac{d\phi}{dt^*} \right)^2 \right]^{1/2}, \tag{V.12b}$$

where

C_D = drag coefficient of the satellite,
G = universal gravitational constant,
M = mass of the earth,
m = mass of the satellite,
n = lift-drag ratio of the satellite,
r^* = radial distance from center of earth to satellite,
S = reference area of the satellite,

([17]) This section is a revised version of the work of Simmons in [22].

t^* = time,
ρ = atmospheric density,
ϕ = true anomaly.

If the following dimensionless variables are defined

$$r = \frac{r^*}{R}; \qquad t = \frac{t^*}{T}; \qquad \epsilon = \frac{C_D \rho S R}{2m}$$

where

R = initial radial distance of the satellite,
T = characteristic time scale = $(R^3/GM)^{1/2}$,
ϵ = aerodynamic parameter (ratio of drag to centrifugal force at $r = R$).

Equations (V.12) may be rewritten in the nondimensional form:

$$\ddot{r} - r\dot{\phi}^2 = -\frac{1}{r^2} - \epsilon[\dot{r} - nr\dot{\phi}][\dot{r}^2 + r^2\dot{\phi}^2]^{1/2}, \tag{V.13a}$$

$$r\ddot{\phi} + 2\dot{r}\dot{\phi} = -\epsilon[r\dot{\phi} + n\dot{r}][\dot{r}^2 + r^2\dot{\phi}^2]^{1/2}. \tag{V.13b}$$

With the **F** implicit in (V.13) we obtain the following values of $\overline{Q_r}$ and $\overline{Q_\phi}$ from (V.2) and (V.7)

$$\overline{Q_r} = p^2(u' + nu)[(u')^2 + u^2]^{1/2}, \tag{V.14a}$$

$$\overline{Q_\phi} = \frac{p^2}{u^2}(nu' - u)[(u')^2 + u^2]^{1/2}. \tag{V.14b}$$

Hence, use of u and t as dependent variables (cf. (V.11)) transforms (V.14) to

$$u'' + u = u^4(t')^2 - \epsilon\frac{n}{u^3}[(u')^2 + u^2]^{3/2}, \tag{V.15a}$$

$$u^2t'' + 2uu't' = \epsilon t'\left[1 - \frac{n}{u}u'\right][(u')^2 + u^2]^{1/2}. \tag{V.15b}$$

The following initial conditions are adopted at $\phi = 0$:

$$u = [a_0(1 + e_0)]^{-1} = 1, \tag{V.16a}$$

$$t = 0, \tag{V.16b}$$

$$u' = 0, \tag{V.16c}$$

$$t' = a_0^{3/2}(1 + e_0^2)(1 - e_0^2)^{-1/2}. \tag{V.16d}$$

For $\epsilon = 0$, the above initial conditions yield a Kepler ellipse with

the following elements:

$a_0 =$ semimajor axis,

$e_0 =$ eccentricity,

$\phi_0 = 0 =$ longitude of apogee,

$\tau_0 = 0 =$ time of apogee passage.

We assume that u and t have the following expansions:

$$u(\phi;\epsilon) = \sum_{n=0}^{N} U^{(n)}(\phi,\tilde{\phi})\,\epsilon^n + O(\epsilon^{N+1}), \tag{V.17a}$$

$$t(\phi;\epsilon) = \frac{T(\tilde{\phi})}{\epsilon} + \sum_{n=0}^{N} T^{(n)}(\phi,\tilde{\phi})\,\epsilon^n + O(\epsilon^{N+1}) \tag{V.17b}$$

where as usual $\tilde{\phi} = \epsilon\phi$.

The need for the term $T(\tilde{\phi})/\epsilon$ in the expansion of t was pointed out by J. A. Morrison in a private communication, and its use will become evident later on.

The following equations for $U^{(0)}$, $U^{(1)}$, $T^{(0)}$ and $T^{(1)}$ now follow from (V.15) and (V.17)

$$U_{11}^{(0)} + U^{(0)} = U^{(0)^4}(T_1^{(0)} + T_2)^2, \tag{V.18a}$$

$$U^{(0)^2}T_{11}^{(0)} + 2U^{(0)}U_1^{(0)}(T_1^{(0)} + T_2) = 0, \tag{V.18b}$$

$$\begin{aligned} U_{11}^{(1)} + U^{(1)} = &-2U_{12}^{(0)} + 2U^{(0)^4}(T_1^{(0)} + T_2)(T_1^{(1)} + T_2^{(0)}) \\ &+ 4U^{(0)^3}U^{(1)}(T_1^{(0)} + T_2)^2 - \frac{n}{U^{(0)^3}}(U_1^{(0)^2} + U^{(0)^2})^{3/2}, \end{aligned} \tag{V.18c}$$

$$\begin{aligned} &U^{(0)^2}(T_{11}^{(1)} + 2T_{12}^{(0)} + T_{22}) + 2U^{(0)}U^{(1)}T_{11}^{(0)} + 2U^{(1)}U_1^{(0)}(T_1^{(0)} + T_2) \\ &+ 2U^{(0)}[U_1^{(0)}(T_1^{(1)} + T_2^{(0)}) + (T_1^{(0)} + T_2)(U_1^{(1)} + U_2^{(0)})] \\ &\qquad = (T_1^{(0)} + T_2)\left(1 - n\frac{U_1^{(0)}}{U^{(0)}}\right)(U_1^{(0)^2} + U^{(0)^2})^{1/2}. \end{aligned} \tag{V.18d}$$

The solution of (V.18a) and (V.18b) is

$$U^{(0)} = l^2[1 - e\cos(\phi - \beta)], \tag{V.19a}$$

$$T^{(0)} = a^{3/2}[\Phi + e\sin\Phi] + \tau - T_2\phi, \tag{V.19b}$$

where

$$\sin\Phi = [(1 - e^2)^{1/2}\sin(\phi - \beta)]/[1 - e\cos(\phi - \beta)], \tag{V.19c}$$

$$l^2 = a(1 - e^2). \tag{V.19d}$$

In (V.19) a, e, β, τ are functions of $\tilde{\phi}$ which according to (V.16) have the initial values a_0, e_0, 0, 0 respectively.

For simplicity we shall henceforth neglect terms of order e^2. This approximation will be justified later by the fact that $e_0 \to 0$ as $\tilde{\phi} \to \infty$.

Equation (V.18d) can be integrated to

$$U^{(0)2}(T_1^{(1)} + T_2^{(0)}) + 2l\frac{U^{(1)}}{U^{(0)}} + \frac{dl}{d\tilde{\phi}}\phi - \frac{1}{l}[\phi + e\sin(\phi - \beta) + ne\cos(\phi - \beta) + O(e^2)] = g(\tilde{\phi}) \tag{V.20}$$

and use of (V.20) in (V.18c) leads to

$$U_{11}^{(1)} + U^{(1)} = \left(2e - 2l^2\frac{de}{d\tilde{\phi}} - 4le\frac{dl}{d\phi}\right)\sin(\phi - \beta) + 2e\left(l^2\frac{d\beta}{d\tilde{\phi}} + n\right)\cos(\phi - \beta) + 2\left(1 - l\frac{dl}{d\tilde{\phi}}\right)\phi + 2lg - n. \tag{V.21}$$

The boundedness of $U^{(1)}$ requires that

$$1 - l\frac{dl}{d\tilde{\phi}} = 0, \tag{V.21a}$$

$$e - l^2\frac{de}{d\tilde{\phi}} - 2le\frac{dl}{d\tilde{\phi}} = 0, \tag{V.21b}$$

$$l^2\frac{d\beta}{d\tilde{\phi}} + n = 0, \tag{V.21c}$$

whose solution defines a, e, and β in the form

$$a = l^{-2} + O(e^2) = (1 + 2\tilde{\phi})^{-1}, \tag{V.22a}$$

$$e = e_0(1 + 2\tilde{\phi})^{-1/2}, \tag{V.22b}$$

$$\beta = -\frac{n}{2}\log(1 + 2\tilde{\phi}). \tag{V.22c}$$

Equation (V.21) may now be solved in the form

$$U^{(1)} = A(\tilde{\phi})\sin(\phi - \beta) + B(\tilde{\phi})\cos(\phi - \beta) + 2lg - n \tag{V.23}$$

where A and B are functions which are to be determined by the boundedness of $U^{(2)}$.

Using (V.23) in (V.20) leads to

$$
\begin{aligned}
T_1^{(1)} = & -\left(\frac{d\tau}{d\tilde{\phi}} + 3a^2 g + 3a^{5/2} eB + 3a^{5/2}\beta - na^{5/2}\right) \\
& + (3a^{5/2} + T_{22})\phi + a^{5/2}(9e - 2A)\sin(\phi - \beta) \\
& - a^{5/2}(2B - 5en + 10a^{-1/2}eg)\cos(\phi - \beta) \\
& - 3a^{5/2} e\,[A \sin 2(\phi - \beta) + B\cos 2(\phi - \beta)\,].
\end{aligned}
\tag{V.24}
$$

Upon integration of (V.24), the first two terms on the right-hand side would lead to contributions proportional to ϕ and ϕ^2 respectively in $T^{(1)}$. This clearly contradicts the consistency of the two variable representation of t unless the terms in question are to vanish identically. One then obtains the following equations defining τ and T,

$$
\frac{d\tau}{d\tilde{\phi}} = -3a^{5/2} eB - 3a^2 g - 3a^{5/2}\beta + a^{5/2} n, \tag{V.25a}
$$

$$
\frac{d^2 T}{d\tilde{\phi}^2} + 3a^{5/2} = 0. \tag{V.25b}
$$

The solution of (V.25) hinges upon the prior determination of g and B, both of which can only be defined by conditions upon the terms of $O(\epsilon^2)$ of the solution.

This property of the time-history of the motion is not a mere consequence of the present expansion of t as in (V.17). Had we adopted the formulation of (V.10) and computed the pair of variables u and p first, the quadrature of (V.10c) leading from the angular momentum p to the time t would have led to the same conclusion: that knowledge of u and p to $O(\epsilon)$ was necessary to compute t to $O(\epsilon)$. This is easily seen by noting that the expansion for t' will in general contain long-period terms proportional to $\frac{\sin}{\cos}(\epsilon\phi)$ which upon integration drop in order by one power of ϵ. Thus, the long-period terms of order ϵ^{n+1} in the expression for t' must be known together with all the terms to order ϵ^n before one can also compute t to $O(\epsilon^n)$ for any n.

If, in contrast to this example, there existed an exact integral of motion (such as the energy integral, for instance), one could use this integral to compute t to any order in ϵ^n from knowledge of u and p only to the same order ϵ^n. Briefly, this comes about because

the integral provides the unknown long-period terms of t' to $O(\epsilon^{n+1})$ once all the other terms in this quantity have been computed to $O(\epsilon^n)$. For a detailed discussion of this point, the reader is referred to [7] and [8].

We now return to the determination of T which follows from (V.25b) and the initial conditions $T(0) = 0$, $T_2(0) = 0$ derived from (V.16b) and (V.16d). One calculates

$$T = (1 - \tilde{\phi}) - (1 + 2\tilde{\phi})^{-1/2}. \tag{V.26}$$

We note first that as $\epsilon \to 0$, $\epsilon^{-1}T \to 0$ as one would expect. Moreover, the role of T in the expansion (V.17b) is clear if we observe that omission of T in (V.17b) will lead to a value of $T^{(1)}$ containing the inconsistent term $(3/2)\,a^{5/2}\phi^2 = (3/2)\,\phi^2 + O(e^2)$ according to (V.24). This is the leading term in the nonuniform representation of T for large values of $\tilde{\phi}$, obtained by expanding (V.26) in its Taylor series around the origin.

The procedure for calculating the higher order terms of this solution essentially parallels that used in prior examples and will not be given.

Although the time-history is only partially defined, the results of (V.22) establish the dominant behavior of the orbit in space. The effect of drag is to cause both the semimajor axis and eccentricity to decay monotonically. Since e continually decreases as the orbit evolves, the omission of the nonlinear terms in e is justified for sufficiently small initial eccentricities. To this order, the only effect of the lift force is a slow regression of the apse, and a net decrease of $U^{(1)}$ by the value n, (cf. (V.23)).

c. *The Motion of a Close Satellite in the Restricted Three-Body Problem.* The author in [15] has shown that for sufficiently close satellites of the smaller primary the following are a limiting form of the equations of the restricted three-body problem:

$$\ddot{x} = -\frac{x}{r^3} + 2\epsilon\dot{y} + 3\epsilon^2 x + O(\epsilon^4), \tag{V.27a}$$

$$\ddot{y} = -\frac{y}{r^3} - 2\epsilon\dot{x} + O(\epsilon^4), \tag{V.27b}$$

$$r^2 = x^2 + y^2. \tag{V.27c}$$

The Cartesian coordinates x, y locate the particle of negligible

mass in a uniformly rotating coordinate system centered around the smaller primary. The ratio of the mass of the smaller primary to that of the system is denoted by $\mu = \epsilon^4$ and the distance and time variables in (V.27) are of the order ϵ^2 and ϵ respectively in terms of the conventional dimensionless variables of the restricted problem.

Even though the x, y system is noninertial we may regard the right-hand sides of (V.27) as the x and y components of $\mathbf{F}$ to compute the following expressions for the generalized perturbation forces (with p and u as in (V.5e) and (V.5f)):

$$\overline{Q_r} = 2pu + \frac{3\epsilon}{u}\cos^2\phi + O(\epsilon^3), \tag{V.28a}$$

$$\overline{Q_\phi} = \frac{2p}{u}\frac{du}{d\tilde{\phi}} - \frac{3\epsilon}{u^2}\sin\phi\cos\phi + O(\epsilon^3). \tag{V.28b}$$

Equations (V.11) then become

$$u'' + u - u^4(t')^2 = -2\epsilon t'\left[u + \frac{1}{u}(u')^2\right] + \frac{3}{2}\epsilon^2(t')^2[u'\sin 2\phi - u(1+\cos 2\phi)] + O(\epsilon^4), \tag{V.29a}$$

$$ut'' + 2u't' = -2\epsilon u'(t')^2 + \frac{3}{2}\epsilon^2 u(t')^3\sin 2\phi + O(\epsilon^4). \tag{V.29b}$$

The author has solved the above by a technique analogous to the method of Lindstedt for the following set of initial conditions:

$$u(0) = a_0^{-1}(1+e_0)^{-1}, \tag{V.30a}$$

$$u'(0) = 0, \tag{V.30b}$$

$$t(0) = 0, \tag{V.30c}$$

$$t'(0) = \frac{a_0^{3/2}(1+e_0)^2}{(1-e_0^2)^{1/2}} + \frac{\epsilon a_0^3(1+e_0)^4}{1-e_0^2} + O(\epsilon^2). \tag{V.30d}$$

Here we shall summarize the analysis using the two variable expansion procedure which duplicates the results of the earlier work by a more systematic approach.

We use ϕ as the fast variable and adopt $\tilde{\phi}$ given by:

$$\tilde{\phi} = \epsilon\phi(1 + \alpha_1\epsilon + \alpha_2\epsilon^2 + \cdots) \tag{V.31}$$

as a slow variable. The reason for distorting the slow variable here rather than the fast one as done previously is because the equations of motion contain ϕ directly and it is therefore a natural variable for the problem.

We expand u and t in the form([18])

$$u = \sum_{n=0}^{N} U^{(n)}(\phi, \tilde{\phi})\,\epsilon^n + O(\epsilon^{N+1}), \tag{V.32a}$$

$$t = \sum_{n=0}^{N} T^{(n)}(\phi, \tilde{\phi})\,\epsilon^n + O(\epsilon^{N+1}). \tag{V.32b}$$

The following lengthy formulae follow from (V.29) upon substitution of (V.32) with $N = 2$:

$$U_{11}^{(0)} + U^{(0)} = U^{(0)4} T_1^{(0)2}, \tag{V.33a}$$

$$U^{(0)} T_{11}^{(0)} + 2U_1^{(0)} T_1^{(0)} = 0, \tag{V.33b}$$

$$\begin{aligned} U_{11}^{(1)} + U^{(1)} = &- 2U_{12}^{(0)} + 2U^{(0)4} T_1^{(0)}(T_1^{(1)} + T_2^{(0)}) \\ &+ 4U^{(0)3} T_1^{(0)2} U^{(1)} - 2U^{(0)} T_1^{(0)} - \frac{2}{U^{(0)}}\, U_1^{(0)2} T_1^{(0)}, \end{aligned} \tag{V.33c}$$

$$\begin{aligned} U^{(0)}(T_{11}^{(1)} + 2T_{12}^{(0)}) + U^{(1)} T_{11}^{(0)} + 2T_1^{(0)}(U_1^{(1)} + U_2^{(0)}) \\ + 2U_1^{(0)}(T_1^{(1)} + T_2^{(0)}) = - 2U_1^{(0)} T_1^{(0)2}, \end{aligned} \tag{V.33d}$$

$$\begin{aligned} U_{11}^{(2)} + U^{(2)} = &- 2U_{12}^{(1)} - U_{22}^{(0)} - 2\alpha_1 U_{12}^{(1)} \\ &+ 2U^{(0)4} T_1^{(0)}(T_1^{(2)} + T_2^{(1)} + \alpha_1 T_2^{(0)}) \\ &+ U^{(0)4}(T_1^{(1)} + T_2^{(0)})^2 + 8U^{(0)3} U^{(1)} T_1^{(0)}(T_1^{(1)} + T_2^{(0)}) \\ &+ 4U^{(0)3} U^{(2)} T_1^{(0)2} + 6U^{(0)2} T_1^{(0)2} U^{(1)2} - 2U^{(0)}(T_1^{(1)} + T_2^{(0)}) \\ &- 2U^{(1)} T_1^{(0)} - \frac{4}{U^{(0)}}\, U_1^{(0)} T_1^{(0)}(U_1^{(1)} + U_2^{(0)}) \\ &- \frac{2}{U^{(0)}}\, U_1^{(0)2}(T_1^{(1)} + T_2^{(0)}) + \frac{2}{U^{(0)2}}\, U_1^{(0)2} T_1^{(0)} U^{(1)} - \frac{3}{2}\, U^{(0)} T_1^{(0)2} \\ &+ \frac{3}{2}\, T_1^{(0)2}(U_1^{(0)} \sin 2\phi - U^{(0)} \cos 2\phi), \end{aligned} \tag{V.33e}$$

([18]) Note that here it is unnecessary to include the term $T(\tilde{\phi})/\epsilon$ in (V.32b) as was done in (V.17b) since the expansion for t will not contain a term proportional to ϕ^2.

$$
\begin{aligned}
&U^{(0)}(T_{11}^{(2)} + 2T_{12}^{(1)} + T_{22}^{(0)} + 2\alpha_1 T_{12}^{(0)}) + U^{(1)}(T_{11}^{(1)} + 2T_{12}^{(0)}) \\
&\quad + U^{(2)}T_{11}^{(0)} + 2T_1^{(0)}(U_1^{(2)} + U_2^{(1)} + \alpha_1 U_2^{(1)}) \\
&\quad + 2U_1^{(0)}(T_1^{(2)} + T_2^{(1)} + \alpha_1 T_2^{(0)}) + 2T_1^{(0)2}(U_1^{(1)} + U_2^{(0)}) \\
&\quad + 2(U_1^{(1)} + U_2^{(0)} + 2U_1^{(0)}T_1^{(0)})(T_1^{(1)} + T_2^{(0)}) \\
&= \frac{3}{2}\, U^{(0)}T_1^{(0)3}\sin 2\phi.
\end{aligned}
\tag{V.33f}
$$

Equations (V.19) with $T = 0$ also define the solutions of (V.33a) and (V.33b), and substitution of these results into (V.33d) leads to the integral

$$
U^{(0)2}(T_1^{(1)} + T_2^{(0)}) + \frac{2}{U^{(0)}}\, lU^{(1)} + \frac{l^2}{U^{(0)2}} - \frac{dl}{d\tilde{\phi}}\,\phi = g(\tilde{\phi}). \tag{V.34}
$$

We again assume e is small([19]) and neglect all nonlinear contributions in the eccentricity in calculating (V.33c) as given below

$$
\begin{aligned}
U_{11}^{(1)} + U^{(1)} = &-2\frac{d(l^2e)}{d\tilde{\phi}}\sin(\phi - \beta) + 2e\left(l^2\frac{d\beta}{d\tilde{\phi}} + \frac{1}{l}\right)\cos(\phi - \beta) \\
&- \frac{dl^2}{d\tilde{\phi}}\,\phi + 2lg.
\end{aligned}
\tag{V.35}
$$

Boundedness of $U^{(1)}$ defines l, e and β as follows

$$
l = \text{constant} = l_0 = [a_0(1 - e_0^2)]^{-1/2}, \tag{V.36a}
$$

$$
e = \text{constant} = e_0, \tag{V.36b}
$$

$$
\beta = -\, l_0^{-3}\tilde{\phi}, \tag{V.36c}
$$

and the solution of (V.35) becomes

$$
U^{(1)} = A(\tilde{\phi})\sin(\phi - \beta) + B(\tilde{\phi})\cos(\phi - \beta) + 2l_0 g. \tag{V.37}
$$

Equation (V.34) can now be integrated to give

$$
\begin{aligned}
T^{(1)} = &\,2(e_0a_0^3 - 5e_0a_0^2g - a_0^{5/2}B)\sin(\phi - \beta) \\
&+ 2a_0^{5/2}A\cos(\phi - \beta) - \left(3a_0^2g + 3e_0a_0^{5/2}B + \frac{d\tau}{d\tilde{\phi}}\right)\phi \\
&+ \frac{3}{2}\, e_0a_0^{5/2}[A\cos 2(\phi - \beta) - B\sin 2(\phi - \beta)] + h(\tilde{\phi}),
\end{aligned}
\tag{V.38}
$$

([19]) The analysis for arbitrary values of e is given by Eckstein in [9].

and we require the term which is linear in ϕ to vanish by setting

$$\frac{d\tau}{d\tilde{\phi}} = -3(ga_0^2 + e_0 a_0^{5/2} B). \tag{V.39}$$

We next integrate (V.33f) in the form

$$\begin{aligned} &U^{(0)2}(T_1^{(2)} + T_2^{(1)} + \alpha_1 T_2^{(0)}) + 2\frac{l_0}{U^{(0)}}\, U^{(2)} - 3\frac{l_0}{U^{(0)2}}\, U^{(1)2} \\ &+ 4\frac{l_0^2}{U^{(0)3}}\, U^{(1)} + 2\frac{g}{U^{(0)}}\, U^{(1)} - \frac{l_0^3}{U^{(0)4}} - 2\frac{l_0}{U^{(0)2}}\, g \\ &+ \frac{dg}{d\tilde{\phi}}\,\phi - Q = k(\tilde{\phi}) \end{aligned} \tag{V.40a}$$

where

$$Q = -\frac{3e_0}{l_0^5}\cos(\phi + \beta) - \frac{3}{4l_0^5}\cos 2\phi - \frac{e_0}{l_0^5}\cos(3\phi - \beta). \tag{V.40b}$$

After some manipulation (V.33c) becomes

$$\begin{aligned} U_{11}^{(2)} + U^{(2)} = &-2l_0\frac{dg}{d\tilde{\phi}}\,\phi \\ &+\left[2\frac{dB}{d\tilde{\phi}}\cos\beta - \left(2\frac{dA}{d\tilde{\phi}} - 8e_0a_0^2 + 6e_0a_0g + 2\alpha_1 e_0 a_0^{1/2}\right)\sin\beta\right]\sin\phi \\ &-\left[2\frac{dB}{d\tilde{\phi}}\sin\beta + \left(2\frac{dA}{d\tilde{\phi}} + 7e_0a_0^2 + 6e_0a_0g + 2\alpha_1 e_0 a_0^{1/2}\right)\cos\beta\right]\cos\phi \\ &- 5e_0a_0\cos(3\phi - \beta) - 3a_0^2\cos 2\phi \\ &- 6e_0a_0^{3/2}[A\sin 2(\phi - \beta) + B\cos 2(\phi - \beta)] \\ &+ 2l_0k + g^2 - \frac{a_0^2}{2} - 2Be_0a_0^{3/2} \end{aligned} \tag{V.41}$$

and elimination of the terms proportional to ϕ, $\sin(\phi - \beta)$ and $\cos(\phi - \beta)$ on the right-hand side of (V.41) defines $g(\tilde{\phi})$, $A(\tilde{\phi})$ and $B(\tilde{\phi})$ as follows

$$g(\tilde{\phi}) = \text{const.} = g_0 = g(0), \tag{V.42a}$$

$$A(\tilde{\phi}) = e_0 \left(\frac{a_0^2}{4} - 3g_0 a_0 - \alpha_1 a_0^{1/2} \right) \tilde{\phi} + \frac{15}{8} e_0 a_0^{1/2} \sin 2\beta + A(0), \tag{V.42b}$$

$$B(\tilde{\phi}) = \frac{15}{8} e_0 a_0^{1/2} (1 - \cos 2\beta) + B(0). \tag{V.42c}$$

The term proportional to $\tilde{\phi}$ in (V.42b) must be eliminated in order that the expansion be consistent, and this defines α_1 as follows

$$\alpha_1 = a_0^{1/2} \left(\frac{a_0}{4} - 3g_0 \right) . \tag{V.43}$$

The initial conditions (V.30) imply that

$$g_0 = A(0) = B(0) = \tau(0) = 0. \tag{V.44}$$

Hence

$$\tau = O(e^2) \tag{V.45}$$

and the results to $O(\epsilon)$ are complete except for $h(\tilde{\phi})$.

A straightforward calculation shows that the above agrees exactly with the author's results in [**15**].

After a slightly more complicated transformation of the various variables one can show that these results are also identical with the "lunar theory" of de Pontécoulant (cf. [**4**]) for the case $i = 0$ and to the order presently computed.

References

1. D. J. Benney, *The flow induced by a disk oscillating in its own plane,* J. Fluid Mech. **18** (1964), 385-391.

2. N. N. Bogoliubov and Y. A. Mitropolsky, *Asymptotic methods in the theory of non-linear oscillations,* Hindustan Publishing Corp., Delhi, India, 1961.

3. F. P. Bretherton, *Resonant interactions between waves: The case of discrete oscillations,* J. Fluid Mech. **20** (1964), 457-479.

4. E. W. Brown, *An introductory treatise on the lunar theory,* Cambridge Univ. Press, New York, 1896.

5. J. D. Cole and J. Kevorkian, *Uniformly valid asymptotic approximations for certain non-linear difierential equations,* Proc. Internat. Sympos. Non-linear Differential Equations and Non-linear Mechanics, Academic Press, New York, 1963, pp. 113-120.

6. M. C. Eckstein, R. R. Shi and J. Kevorkian, *Satellite motion for all inclinations around an oblate planet,* Proc. I.A.U. Sympos. No. 25, 1964, Thessaloniki, Greece. (to be published)

7. _____, *Satellite motion for arbitrary eccentricity and inclination around the smaller primary in the restricted three-body problem,* A.I.A.A. Reprint No. 65-690, 1965. (to be published in Astronom. J. **71** (1966).

8. _____, *Time history of a satellite of an oblate planet,* Paper No. 3529, Douglas Aircraft Co., Inc., Santa Monica, California, 1965.

9. M. C. Eckstein, *Application of the two variable method to the motion of a satellite in the restricted three-body problem for zero inclination and arbitrary eccentricity.* (to be published).

10. A. Erdélyi, *Asymptotic expansions,* Dover, New York, 1956.

11. C. S. Gardner, *Adiabatic invariants of periodic classical systems,* Phys. Rev. **115** (1959), 791-794.

12. S. Kaplun and P. A. Lagerstrom, *Asymptotic expansions of Navier-Stokes solutions for small Reynolds numbers,* J. Math. Mech. **6** (1957), 585-593.

13. J. Kevorkian, *The uniformly valid asymptotic representation of the solutions of certain non-linear ordinary differential equations,* Ph.D. Thesis, California Institute of Technology, Pasadena, California, 1961.

14. _____, *The two variable expansion procedure for the approximate solution of certain non-linear differential equations,* Douglas Aircraft Co., Inc., Santa Monica, California, Report No. SM-42620, 1962.

15. _____, *Uniformly valid asymptotic representation for all times of the motion of a satellite in the vicinity of the smaller body in the restricted three-body problem,* Astronom. J. **67** (1962), 204-211.

16. N. M. Krylov and N. N. Bogoliubov, *Introduction to nonlinear mechanics,* Acad. Sci. Ukrain. S.S.R. (1937); translated by S. Lefschetz, Princeton Univ. Press, Princeton, N. J., 1947.

17. G. N. Kuzmak, *Asymptotic solutions of non-linear second order differential equations with variable coefficients,* Prikl. Mat. Meh. **13** (1959), 515-526.

18. J. E. McCune, *Master equation for plasmas,* Phys. Fluids **7** (1964), 1306-1320.

19. N. W. McLachlan, *Theory and application of Mathieu functions,* Clarendon Press, Oxford, 1947.

20. J. A. Morrison, *Comparison of the modified method of averaging and the two variable expansion procedure,* SIAM Rev. **8** (1966), 66-85.

21. H. Poincaré, *Les méthodes nouvelles de la mécanique céleste,* Vol. II, Dover, New York, 1957.

22. J. L. Simmons, *Application of the two variable expansion procedure to problems in celestial mechanics,* Engineer's Thesis, California Institute of Technology, Pasadena, 1962.

23. J. J. Stoker, *Non-linear vibrations in mechanical and electrical systems,* Interscience, New York, 1950.

24. R. A. Struble, *A geometrical derivation of the satellite equations,* J. Math. Anal. Appl. **1** (1960), 300.

25. _____, *Non-linear differential equations,* McGraw-Hill, New York, 1962.

26. _____, *The geometry of the orbits of artificial satellites,* Arch. Rational Mech. Anal. **7** (1961), 87.

UNIVERSITY OF WASHINGTON,
SEATTLE, WASHINGTON

J. C. Houbolt

Rendezvous Problem

I. **Basic orbital information.** The definitions of the major symbols are:

a: semimajor axis
E: eccentric anomaly
V: true anomaly
p: parameter of ellipse—semilatus rectum
T: period
i: inclination angle
ω: argument of perigee
Ω: argument of ascending node
e: eccentricity
R: equatorial radius of Earth
r: distance from Earth's center
τ: time at epoch.

In terms of these quantities we will use the well-known relations:

(I.1) $r_p = a(1 - e)$; pericenter distance

(I.2) $r_a = a(1 + e)$; apocenter distance

(I.3) $p = a(1 - e^2)$; semilatus rectum

(I.4) $T = 2\pi a^{3/2}\mu^{-1/2}$; period

μ: $1.407639 \times 10^{16}\,\mathrm{ft}^3/\mathrm{sec}^2$ for the Earth's gravitational constant

$$n = \bar{\omega} = \frac{2\pi}{T} = \mu^{1/2} a^{-3/2}; \quad \text{mean motion} \tag{I.5}$$

$$M = n(t - \tau); \quad \text{mean anomaly.} \tag{I.6}$$

If the Earth's potential function is represented by

$$U = \frac{\mu}{r}\left[1 - \sum_{n=2}^{\infty} J_n \left(\frac{R}{r}\right)^n P_n(\sin L)\right] \tag{I.7}$$

where

P_n = Legendre polynomial of order n,

L = latitude angle,

then the first order secular perturbations in the orbital elements of an Earth satellite in the absence of air drag are:

$$\dot{\Omega}_s = -\frac{3}{2a}\left(\frac{\mu}{a}\right)^{1/2} J_2 \left(\frac{R}{p}\right)^2 \cos i \,\mathrm{rad/sec} \tag{I.8}$$

$$\dot{\omega}_s = \frac{3}{4a}\left(\frac{\mu}{a}\right)^{1/2} J_2 \left(\frac{R}{p}\right)^2 (-1 + 5\cos^2 i)\,\mathrm{rad/sec} \tag{I.9}$$

$$\dot{M}_s = \frac{3}{4a}\left(\frac{\mu}{a}\right)^{1/2} J_2 \left(\frac{R}{p}\right)^2 (1 - e^2)^{1/2}(-1 + 3\cos^2 i)\,\mathrm{rad/sec} \tag{I.10}$$

$$\dot{\Omega}_s = -3\pi J_2 \left(\frac{R}{p}\right)^2 \cos i \,\mathrm{rad/rev} \tag{I.11}$$

$$\dot{\omega}_s = 3\pi J_2 \left(\frac{R}{p}\right)^2 (-1/2 + 5/2\cos^2 i)\,\mathrm{rad/rev} \tag{I.12}$$

$$\dot{M}_s = 3\pi J_2 \left(\frac{R}{p}\right)^2 (1 - e^2)^{1/2}(-1/4 + 3/4\cos^2 i)\,\mathrm{rad/rev} \tag{I.13}$$

where $J_2 = 1082.28 \times 10^{-6}$.

EXAMPLE. For an orbit with $i = 30°$ and an altitude of 300 statute miles, one finds that

$$\dot{\Omega} = -0.442°/\mathrm{rev} \approx -6.8°/\mathrm{day},$$

$$\dot{\omega} = 0.705°/\mathrm{rev} \approx 10.8°/\mathrm{day}.$$

II. **Rendezvous phases.** Rendezvous can be divided into the three phases:

(i) ascent and injection into transfer orbit

(ii) terminal phase

(iii) docking—contact between ferry and target vehicles.

In the present chapter, we will give no discussion of (iii).

There are a wide variety of possible types of ascent maneuvers and a few remarks will be made concerning the characteristics of some of the basic types of ascent maneuvers.

a. *In-plane Ascent.* For this, the launch site must be in the orbital plane of the target.

An in-plane ascent requires that the target vehicle travel in a compatible orbit; that is, an orbit in which the target passes over the launch site at least once per day. This is a severe requirement and its practical realization will probably require means for adjusting the orbital plane of the target vehicle.

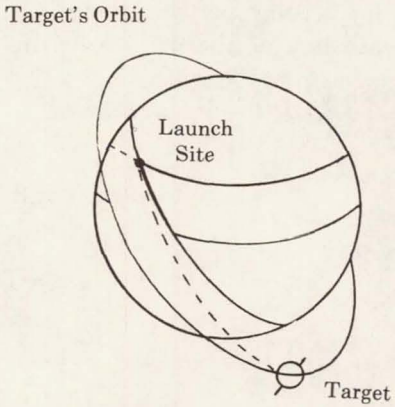


FIGURE 1. Transfer Orbit

b. *Adjacency Transfer.* The ferry is inserted into an orbit close to that of the target, but not necessarily in the same orbital plane. The ferry transfer orbit is selected so that its orbit is at the same altitude and has the same speed as the target at the time at which the two objects intersect. At the time of intersection, the ferry is given a velocity impulse such that its orbit plane is made coincident with that of the target.

c. *Two-impulse Transfer.* The first impulse inserts the ferry into a transfer orbit such that the apogee of the transfer occurs at the orbit of the ferry and the timing is such that the ferry and target

are simultaneously at the apogee of the transfer orbit. When the two orbits touch, a second velocity impulse is given to the ferry to bring it up to orbital speed and, if necessary, change its orbital plane to coincide with that of the target.

d. *General Ascent.* The ferry is injected into a general transfer orbit which is required to intersect the target on either the outgoing leg or the incoming leg. The timing problem for these ascents is very critical and typical launch windows are only of about 3 minutes in duration.

e. *Parking Orbits.* An intermediate parking orbit greatly simplifies the timing problems for an ascent transfer trajector. The ferry is first launched into a circular orbit at a lower altitude than that of the target. Because the ferry will have a shorter period of revolution, it will gain on the target with respect to their geocentric angles. At the proper time, the ferry is given a velocity impulse into a transfer orbit which will bring it into position for the final rendezvous maneuver.

III. **Velocity penalty for maneuvers.**

a. *Equal Velocities, In-plane Maneuvers.* Suppose that the interceptor (ferry) and the target vehicles have the same velocity magnitude but different directions; Figure 2.

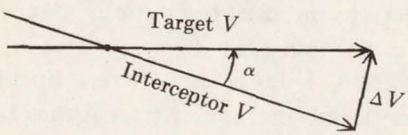


FIGURE 2. In-plane Maneuver

For small α,

$$\Delta V = \alpha V. \tag{III.1}$$

For a typical velocity of 25,000 ft/sec, the velocity increment required per degree separation of the paths would be of the order

$$\Delta V = \frac{\pi}{180} \times 25 \times 10^3 = 436 \text{ ft/sec}. \tag{III.2}$$

This is a costly maneuver as measured in units of required velocity impulse.

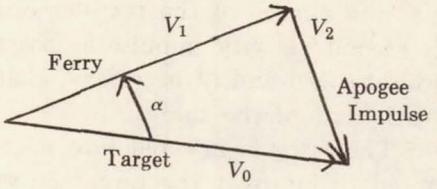


FIGURE 3. Two-impulse Maneuver

b. *Two-impulse Maneuvers.* From Figure 3,

$$V_2^2 = V_1^2 + V_0^2 - 2V_1V_0\cos\alpha. \tag{III.3}$$

The velocity penalty for the plane change is

$$\Delta V = V_1 + V_2 - V_0, \tag{III.4}$$

for small α, such that $\sin\alpha \approx \alpha$, (III.3) and (III.4) yield

$$\Delta V = \frac{V_1V_0}{2(V_0 - V_1)}\alpha^2. \tag{III.5}$$

EXAMPLE. Typical numbers at apogee are $V_1 = 10 \times 10^3$ft/sec, $V_2 = 15 \times 10^3$, $V_0 = 25 \times 10^3$ft/sec. If $\alpha = 5.7°$, then $\Delta V \approx 83$ ft/sec. Thus the two-impulse maneuver is less costly than the previous case. The economy partly comes from the fact that the velocity impulse can correct the interceptor's speed at the same time that its orbital plane is shifted.

c. *Dog-leg Maneuvers.* Dog-leg maneuvers during thrusting may also be used during ascent or transfer trajectories to effect an orbital plane shift. Thrust is made in the transverse direction by tilting the rocket thrust by an angle δ from the vehicle's flight path. Let

ΔV = required increment of velocity.

It can be demonstrated that if $\dot{y} \ll \dot{x}_0$, then for δ held constant,

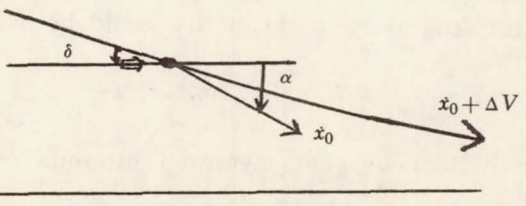


FIGURE 4. Dog-leg Maneuvers

$$\frac{\dot{y}}{\dot{x}_0} = \alpha = \frac{\Delta V}{\dot{x}_0}\,\delta. \tag{III.6}$$

EXAMPLE. If $\dot{x}_0 = 10 \times 10^3$ft/sec, $\Delta V = 15 \times 10^3$ft/sec, one finds that $\alpha = 1.5\delta$.

Thus the dog-leg maneuver can change the angle of the trajectory plane on the same order as the rocket motor gimbel angle used, and with minor penalty on the forward acceleration.

IV. **Ascent.**

a. *General Direct Ascent.* The rendezvous window is defined as the interval of time on the launch pad during which a rendezvous ascent can be made without an "excess" fuel penalty.

It has been established that Hohmann-type transfers require minimum energy (see dashed line in Figure 5). *Soft-rendezvous* is the situation in which the speed and orbit direction are the same for both the interceptor and target vehicles. A Hohmann-type transfer can be used if the target is at A_L at interceptor launch (ahead of insertion point). The intercept takes place at A_R.

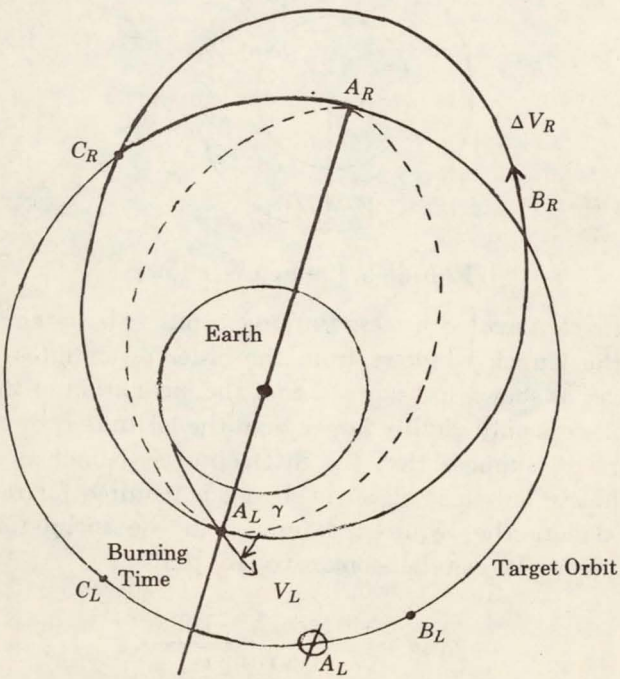


FIGURE 5. Transfer Orbits

The general cases occur for the target at either B_L (leading) or C_L (lagging) with the intercept accomplished at the intersection points B_R or C_R, respectively.

One can investigate the maximum spread in angle between initial points B_L and C_L, which determine the allowable launch window, with a restriction on the available ΔV capability of the interceptor.

Suppose that the total vehicle thrust capability is

$$V_L + \Delta V_R = 27{,}000 \text{ ft/sec.} \tag{IV.1}$$

One can show that the launch window shown in Figure 6 is -7.4 to 6.1° or roughly 13°, which corresponds to about 3 minutes for typical orbits. If the thrust capability is increased to 3×10^4ft/sec, the launch window increases to about 15 minutes. It is thus seen that the launch window is very sensitive to the total vehicle capability.

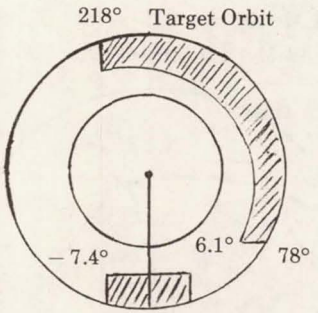


FIGURE 6. Launch Windows

b. *Indirect Ascent Schemes.* Parking orbits can be employed to extend the launch windows from the order of minutes to hours.

Suppose, as shown in Figure 7, that the inclination of the target's orbital plane is only slightly larger than the latitude i_L of the launch site. Further, suppose that the interceptor is launched in a close orbit. That is, only a small angle change is required for rendezvous. Let Δi denote the required difference in the inclination of the orbital plane. It can be demonstrated that

$$\cos\theta = \frac{\sin i_L \cos i_0 - \sin \Delta i}{\cos i_L \sin i_0}. \tag{IV.2}$$

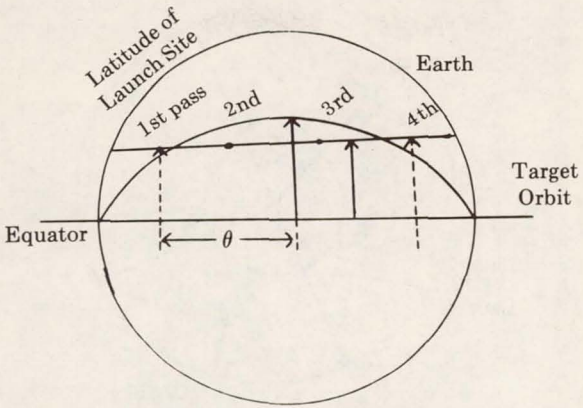


FIGURE 7. Nearly In-plane Launches

EXAMPLE. If $i_L = 28°$ (Cape Kennedy)

i_0	Δ_i	θ
30°	2°	32.6°
30.4	2.4	36.0
31.0	3.0	39.5

Next consider two types of transfer orbits:

Case a. Transfer apogee at target height (Gemini Program Maneuver). A chasing orbit is obtained by launching a transfer such that the apogee is tangent to the target's orbit. Thus the ferry or interceptor gains on the target during each revolution until a constellation is attained for which a single small impulse is sufficient to effect the rendezvous.

Let:

θ = angular difference,

ν = number of revolutions required to overcome θ deficiency.

One can verify that

$$\frac{\Delta r}{r} = \frac{4}{3}\frac{\theta}{360\nu}, \tag{IV.3}$$

$$\frac{\Delta V}{V_0} = \frac{1}{3}\frac{\theta}{360\nu}, \tag{IV.4}$$

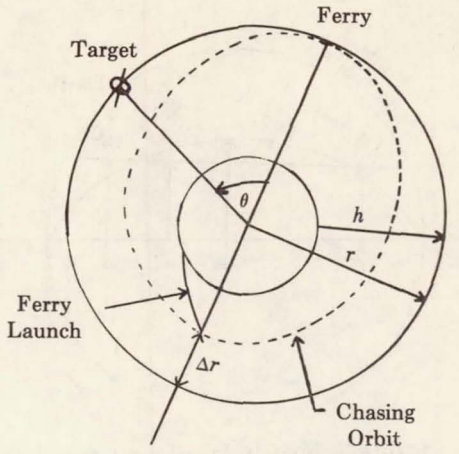


FIGURE 8. Hohmann-type Transfer

where V_0 is the orbital speed.

EXAMPLE. If

$$\theta = 20°,$$

$$\nu = 1,$$

$$r = 4260,$$

then $\Delta r = 315$. This cannot be accomplished in one revolution because Δr is greater than target altitude, here considered to be 300 miles. Therefore, let $\nu = 2$, and then $\Delta r = 158$ miles and $\Delta V = 213$ ft/sec.

Case b. Parking orbit. For an intermediary parking orbit, (IV.3) is modified to

$$\frac{\Delta r}{r} = \frac{2}{3}\frac{\theta}{360(\nu - 3/4)}. \tag{IV.5}$$

Thus the basic technique in the use of chasing or parking orbits is to launch the ferry any time it is ready during the time interval the launch site is close to the orbital plane of the target, Figure 7. From this figure and the table relating Δi and θ, this may be in the interval of several successive orbital passes. Any geocentric angular deficiency that the ferry may have is made up by use of the chasing or parking orbit. It is seen that the holding back for

subsequent addition of a rendezvous velocity increment ΔV_R allows this type rendezvous to be made at substantially the same characteristic velocity increment as would be involved in a direct ascent rendezvous. These indirect schemes provide for launch windows up to 3-5 hours, instead of minutes.

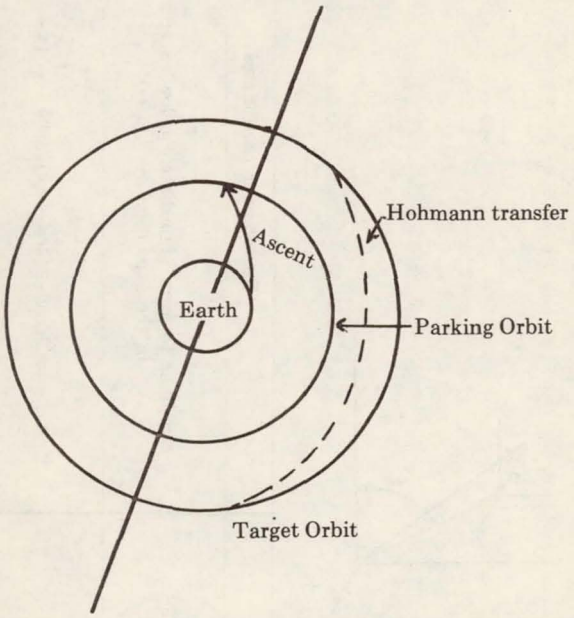


FIGURE 9. Parking Orbit

V. **Terminal phase.** Terminal phase starts when ferry is about 50 miles from the target.

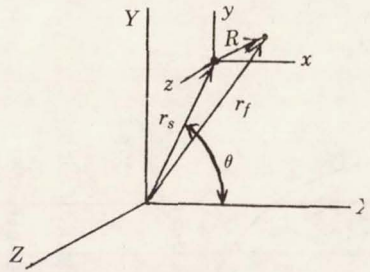
Two types of terminal maneuvers are usually considered.

(i) Proportional navigation: maintain line-of-sight fixed in inertial space; or, maintain zero angular rate.

(ii) Orbital mechanics: compute coast orbits of target and ferry to determine if they intersect. If no intersection, compute required change in ferry orbit to produce orbit intersection.

a. *Terminal Guidance.* The terminal guidance equations for a variety of assumed models are given Tables 1 and 2. Explanation of these tables and much other information can be found in: J. C. Houbolt, *Problems and potentialities of space rendezvous,* Astro-

TABLE I. Terminal Guidance Equations with Inertially Fixed Axes

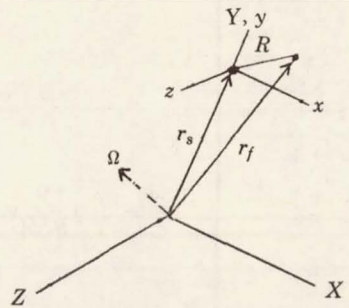


	Vector form	Rectangular coordinates
Exact	$\frac{\partial^2 \bar{r}_f}{\partial t^2} - \nabla \frac{\mu}{r_f} = \frac{\bar{T}}{m}$	Similar to form immediately below except in the expansion of $\nabla (\mu/r_f)$
Spherical earth	$\frac{d^2 \bar{r}_f}{dt^2} + \frac{GM}{r_f^3} \bar{r}_f = \frac{\bar{T}}{m}$ $\bar{r}_f = \bar{r}_s + \bar{R}$ $\bar{r}_s = (X, Y, Z)$ $= (r_s \cos\theta, r_s \sin\theta, 0)$ $\bar{R} = (x, y, z)$	$\ddot{x} + (\ddot{r}_s - r_s\dot{\theta}^2)\cos\theta - (2\dot{r}_s\dot{\theta} + r_s\ddot{\theta})\sin\theta + \frac{GM}{r_f^3}(x + r_s\cos\theta) = \frac{T_x}{m}$ $\ddot{y} + (\ddot{r}_s - r_s\dot{\theta}^2)\sin\theta + (2\dot{r}_s\dot{\theta} + r_s\ddot{\theta})\cos\theta + \frac{GM}{r_f^3}(y + r_s\sin\theta) = \frac{T_y}{m}$ $\ddot{z} + \frac{GM}{r_f^3} z = \frac{T_z}{m}$

TABLE I (continued)

	Vector form	Rectangular coordinates
Spherical earth, Station in a circular orbit	$\frac{d^2\bar{R}}{dt^2} + \frac{d^2\bar{r}_s}{dt^2} + \frac{GM}{r_f^3}\bar{r}_f = \frac{\bar{T}}{m}$	$\ddot{x} - r_s\omega^2\cos\theta + \frac{GM}{r_f^3}(x + r_s\cos\theta) + \frac{T_x}{m}$ $\ddot{y} - r_s\omega^2\sin\theta + \frac{GM}{r_f^3}(y + r_s\sin\theta) = \frac{T_y}{m}$ $\ddot{z} + \frac{GM}{r_f^3}z = \frac{T_z}{m}$
Spherical earth, Circular orbit, 1st order gravity field	$\frac{d^2\bar{R}}{dt^2} + \frac{GM}{r_s^3}\left(\bar{R} - 3\frac{\bar{r}_s \cdot \bar{R}}{r_s^2}\bar{r}_s\right) = \frac{\bar{T}}{m}$	$\ddot{x} + \omega^2(x - 3x\cos^2\theta - 3y\sin\theta\cos\theta) = \frac{T_x}{m}$ $\ddot{y} + \omega^2(y - 3x\sin\theta\cos\theta - 3y\sin^2\theta) = \frac{T_y}{m}$ $\ddot{z} + \omega^2 z = \frac{T_z}{m}$
Spherical earth, Circular orbit, "Zero-order" gravity		
No gravity	$\frac{d^2\bar{R}}{dt^2} = \frac{\bar{T}}{m}$ any orbit	Circular orbit $\ddot{x} = \frac{T_x}{m}$ $\ddot{y} = \frac{T_y}{m}$ $\ddot{z} = \frac{T_z}{m}$

TABLE II. Terminal Guidance Equations with a Rotating Set of Axes



Comments	Vector form	Rectangular coordinates
$\bar{r}_f = \bar{r}_s + \bar{R}$ μ/r_f is the gravity potential due to earth, moon, planets, etc.	$\frac{\partial^2 \bar{r}_f}{\partial t^2} + 2\bar{\Omega} \times \frac{\partial \bar{r}_f}{\partial t} + \dot{\bar{\Omega}} \times \bar{r}_f + \bar{\Omega} \times \bar{\Omega} \times \bar{r}_f - \nabla \frac{\mu}{r_f} = \frac{\bar{T}}{m}$	Similar to form immediately below except in the expansion of $\nabla (\mu/r_f)$
θ is the angular velocity of station about center of earth, r_s is radial position of station	$\frac{d^2 \bar{r}_f}{dt^2} + 2\bar{\Omega} \times \frac{d\bar{r}_f}{dt} + \dot{\bar{\Omega}} \times \bar{r}_f + \bar{\Omega} \times \bar{\Omega} \times \bar{r}_f + \frac{GM}{r_f^3} \bar{r}_f = \frac{\bar{T}}{m}$ $\bar{r}_f = (x, y + r_s, z) = \bar{r}_s + \bar{R}$ $\bar{r}_s = (0, r_s, 0)$ $R = (x, y, z)$ $\bar{\Omega} = (0, 0, \dot{\theta})$	$\ddot{x} - (y + r_s)\ddot{\theta} - 2(\dot{y} + \dot{r}_s)\dot{\theta} - x\left(\dot{\theta}^2 - \frac{GM}{r_f^3}\right) = \frac{T_x}{m}$ $\ddot{y} + x\ddot{\theta} + 2\dot{x}\dot{\theta} + \ddot{r}_s - (y + r_s)\left(\dot{\theta}^2 - \frac{GM}{r_f^3}\right) = \frac{T_y}{m}$ $\ddot{z} + \frac{GM}{r_f^3} z = \frac{T_z}{m}$

TABLE II (continued)

Comments	Vector form	Rectangular coordinates
$r_s = \text{Constant}$ $\dot{\theta} = \text{Constant} = \omega$ $\bar{\Omega} = \text{Constant} = (0,0,\omega)$ $= \bar{\omega}$ $\omega^2 = \frac{GM}{r_s^3} = \dot{\theta}^2$	$\frac{d^2\bar{R}}{dt^2} + 2\bar{\omega}\times\frac{d\bar{R}}{dt} + \bar{\omega}\times\bar{\omega}\times\bar{r}_f + \frac{GM}{r_f^3}\bar{r}_f = \frac{\bar{T}}{m}$ or $\frac{d^2\bar{R}}{dt^2} + 2\bar{\omega}\times\frac{d\bar{R}}{dt} + \omega^2\bar{k}z + \left(\frac{GM}{r_f^3} - \omega^2\right)\bar{r}_f = \frac{\bar{T}}{m}$	$\ddot{x} - 2\omega\dot{y} - x\left(\omega^2 - \frac{GM}{r_f^3}\right) = \frac{T_x}{m}$ $\ddot{y} + 2\omega\dot{x} - (y + r_s)\left(\omega^2 - \frac{GM}{r_f^3}\right) = \frac{T_y}{m}$ $\ddot{z} + \frac{GM}{r_f^3}z = \frac{T_z}{m}$
$\frac{d^2\bar{r}_s}{dt^2} = -\omega^2\bar{r}_s;\ \frac{GM}{r_s^3} = \omega^2$ On the left, $\frac{GM}{r_f^3} = \frac{GM}{r_s^3}\left(1 - 3\frac{\bar{r}_s\cdot\bar{R}}{r_s^2} + \cdots\right)$ $\frac{GM}{r_f^3}\bar{r}_f \cong \omega^2\left(\bar{r}_s + \bar{R} - 3\frac{\bar{r}_s\cdot\bar{R}}{r_s^2}\bar{r}_s\right)$ On the right, $\frac{GM}{r_f^3} = \frac{GM}{r_s^3}\left(1 - 3\frac{y}{r_s} + \cdots\right)$ $\frac{GM}{r_f^3}\bar{r}_f \cong \omega^2(\bar{r}_s + \bar{R} - \bar{j}3y)$	$\frac{d^2\bar{R}}{dt^2} + 2\bar{\omega}\times\frac{d\bar{R}}{dt} + \bar{\omega}\times\bar{\omega}\times\bar{R} + \frac{GM}{r_s^3}(\bar{R} - \bar{j}3y) = \frac{\bar{T}}{m}$ or $\frac{d^2\bar{R}}{dt^2} + 2\bar{\omega}\times\frac{d\bar{R}}{dt} + \frac{GM}{r_s^3}(-\bar{j}3y + \bar{k}z) = \frac{\bar{T}}{m}$	$\ddot{x} - 2\omega\dot{y} = \frac{T_x}{m}$ $\ddot{y} + 2\omega\dot{x} - 3\omega^2 y = \frac{T_y}{m}$ (T1) $\ddot{z} + \omega^2 z = \frac{T_z}{m}$

TABLE II (continued)

Comments	Vector form	Rectangular coordinates
$\frac{GM}{r_f^3}\,\bar{r}_f \cong \frac{GM}{r_s^3}\,(\bar{r}_s + \bar{R})$ $\frac{GM}{r_s^3} = \omega^2$	$\frac{d^2\bar{R}}{dt^2} + 2\bar{\omega}\times\frac{d\bar{R}}{dt} + \bar{\omega}\times\bar{\omega}\times\bar{R} + \frac{GM}{r_s^3}\,\bar{R} = \frac{\bar{T}}{m}$ or $\frac{d^2\bar{R}}{dt^2} + 2\bar{\omega}\times\frac{d\bar{R}}{dt} + \frac{GM}{r_s^3}\,\bar{k}z = \frac{\bar{T}}{m}$	$\ddot{x} - 2\omega\dot{y} = \frac{T_x}{m}$ $\ddot{y} + 2\omega\dot{x} = \frac{T_y}{m}$ $\ddot{z} + \omega^2 z = \frac{T_z}{m}$
	$\frac{d^2\bar{r}_f}{dt^2} + 2\bar{\Omega}\times\frac{d\bar{r}_f}{dt} + \dot{\bar{\Omega}}\times\bar{r}_f + \bar{\Omega}\times\bar{\Omega}\times\bar{r}_f = \frac{\bar{T}}{m}$ for any orbit $\frac{d^2\bar{R}}{dt^2} + 2\bar{\omega}\times\frac{d\bar{R}}{dt} + \bar{\omega}\times\bar{\omega}\times\bar{r}_f = \frac{\bar{T}}{m}$ for circular orbit	Circular orbit $\ddot{x} - 2\omega\dot{y} - \omega^2 x = \frac{T_y}{m}$ $\ddot{y} + 2\omega\dot{x} - \omega^2(y + r_s) = \frac{T_y}{m}$ $\ddot{z} = \frac{T_z}{m}$

nautica Acta, Volume VII, Fasc. 5-6, 1961. As an example, consider the equations in rotating rectangular coordinates for a model having a spherical earth, circular target orbit, "zero-order" gravity. The equations of motion are

$$\ddot{x} - 2\omega\dot{y} = \frac{T_x}{m},$$

$$\ddot{y} + 2\omega\dot{x} - 3\omega^2 y = \frac{T_y}{m}, \tag{V.1}$$

$$\ddot{z} + \omega^2 z = \frac{T_z}{m}.$$

Assume no thrust, $T_x = T_y = T_z = 0$. Then the solutions to (V.1) are

$$x = \left(x_0 + 2\frac{\dot{y}_0}{\omega}\right) + (-3\dot{x}_0 + 6\omega y_0)t - 2\left(3y_0 - 2\frac{\dot{x}_0}{\omega}\right)\sin\omega t - 2\frac{\dot{y}_0}{\omega}\cos\omega t, \tag{V.2}$$

$$y = \left(4y_0 - 2\frac{\dot{x}_0}{\omega}\right) + \left(-3y_0 + 2\frac{\dot{x}_0}{\omega}\right)\cos\omega t + \frac{\dot{y}_0}{\omega}\sin\omega t, \tag{V.3}$$

$$z = a_1 \sin\omega t + b_1 \cos\omega t, \tag{V.4}$$

where x_0, $\dot{x}_0$, y_0, $\dot{y}_0$ are the initial conditions.

The general relative motion seen by the ferry in these coordinates is shown in Figure 10. The ellipse is centered near the target and has the following parameters:

$$v = -3\dot{x}_0 + 6\omega y_0,$$

$$x_c = x_0 + 2\frac{\dot{y}_0}{\omega},$$

$$y_c = 4y_0 - 2\frac{\dot{x}_0}{\omega}, \tag{V.5}$$

$$a = 2b,$$

$$b = \left[\left(\frac{\dot{y}_0}{\omega}\right)^2 + \left(3y_0 - 2\frac{\dot{x}_0}{\omega}\right)^2\right]^{1/2}.$$

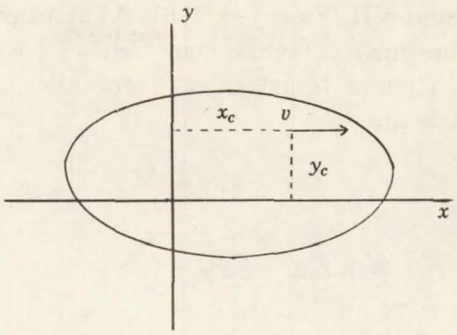


FIGURE 10. Apparent Target Motion

Suppose that $v = y_c = 0$; this implies that

$$\dot{x}_0 = 2\omega y_0, \tag{V.6}$$

which is the condition for which the orbital period of the transfer orbit is equal to that of the target orbit.

If

$$\begin{aligned} \dot{x}_0 &= 2\omega y_0, \\ \dot{y}_0 &= -\frac{\omega x_0}{2}, \end{aligned} \tag{V.7}$$

we have the situation illustrated in Figure 11, in which the ellipse is centered right at the target.

If the target is itself in elliptical motion, it can be shown that the same form of terminal equations apply to the relative motion.

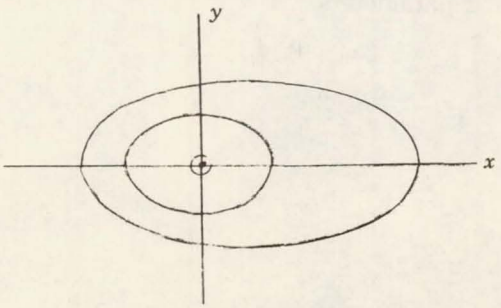


FIGURE 11. Target on Center

b. *Proportional Navigation.* Let R be the line-of-sight between the station and the vehicle. For proportional navigation, an intercept occurs if the direction of the line of sight R in Figure 12 is held inertially fixed in space. One can produce this state of affairs by accelerating the ferry sidewise in the direction of apparent motion of the space station until the space station has no perceptible sidewise motion as viewed against the fixed stars from the ferry. If sidewise motion should reappear, it can be eliminated by a repetition of this maneuver. Other than that, any acceleration of the ferry for completing the rendezvous should be along R.

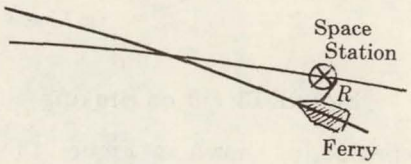


FIGURE 12. Proportional Navigation

First one must accelerate the ferry along R until it is actually approaching the space station. That is, $\dot{R}$ is negative. Ideally, one would now apply a constant acceleration α away from the space station. If α is chosen to satisfy

$$\dot{R}^2 = 2\alpha R \tag{V.8}$$

for the initial values of $\dot{R}$ and R, then (V.8) will continue to hold by the laws of mechanics, and the ferry will come to rest just at the instant of contact with the space station.

In practice, one may not know $\dot{R}$ and R initially to this degree of accuracy, or α may be smaller than can be conveniently maintained by the braking rockets of the ferry. Then the braking is done by an off-on procedure illustrated in Figure 13. The lines marked α_1 and α_2 are ideal approaches with fixed accelerations α_1 and α_2; α_1 is less than α_2 and both are very small. The line for α_1 is called the off line, and that for α_2 the on line. One coasts until it appears the on line has been reached (since this need not be exact, approximate determinations of $\dot{R}$ and R will suffice). Then a thrust is applied toward the space station, producing an acceleration α_3 greater than α_2. This acceleration is maintained until the off line is reached, when the thrust is cut off and coasting is re-

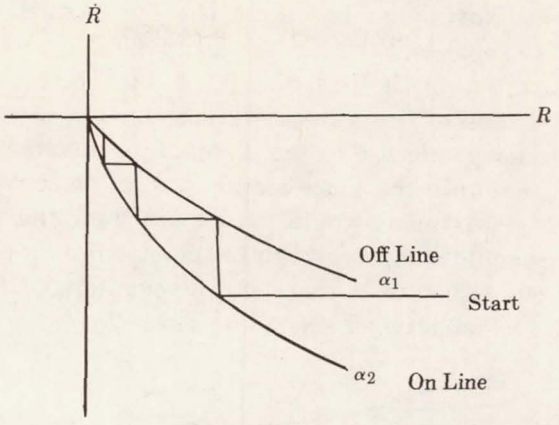


FIGURE 13. Off-on Braking

sumed. This is repeated, as shown in Figure 13 until the ferry is close to the space station and nearly at rest relative to it, and one is ready for docking.

c. *Two Impulse Terminal Scheme.* This is also called the "orbital mechanics scheme." In this scheme, an aiming impulse Δv_1 is first given to direct the ferry on an interception course, and then a correcting impulse Δv_2 is given at rendezvous to prevent a forcible collision (see Figure 14).

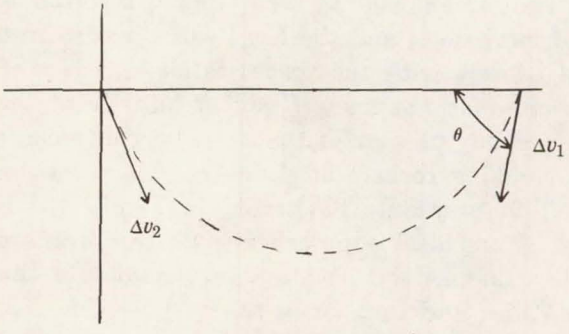


FIGURE 14. Orbital Mechanics

Let $\omega/2\pi$ be the period of the target and t_r be the time interval required to effect a rendezvous. Figure 15 displays the results for the aiming angle θ and the velocity parameter β. If the space station is behind the ferry by a distance x_0, then one fires a rocket to

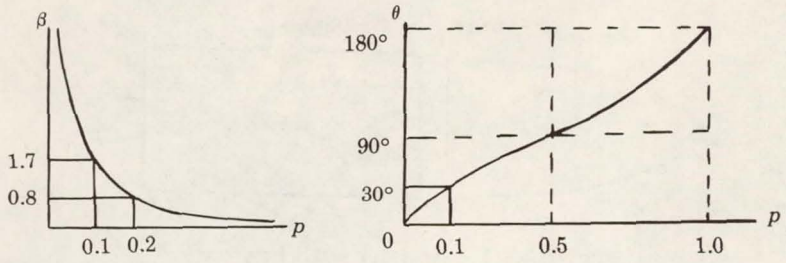


FIGURE 15. Velocity Parameter and Angle

produce a velocity increment

$$\Delta v = \beta \omega x_0 \tag{V.9}$$

in the direction θ, where θ and β are to be read from Figure 15, determining p by the relation

$$p = \frac{\omega}{2\pi} t_r \qquad (t_r \text{ in seconds}). \tag{V.10}$$

EXAMPLE. For an altitude $h = 200$ miles, consider that there is a distance of 5000 feet and a desire to rendezvous in 10 minutes. For this altitude, $\omega = 0.00114$, so that $\Delta v = 5.7\beta$ and $p = 0.109$ by (V.9) and (V.10). Then by Figure 15, we read $\beta = 1.6$ and $\theta = 35°$. So a velocity increment of 9.1 ft/sec, directed 35° below the line of sight is called for. (We interpret "below" as toward the earth's center.) If the ferry were 5000 ft in front of the space station, it would be given an impulse of 9.1 ft/sec backward and upward 35°.

VI. **Mission analysis.** Mission analysis is used for booster design, or specifying the rocket thrust capabilities. As an example of mission design, consider a comparison of two types of lunar mission profiles:

(i) direct ascent;

(ii) direct ascent with rendezvous in parking orbit about moon.

a. *Basic Rocket Equations.* The basic rocket equation can be written as

$$\frac{m_0}{m} = K = e^{(\Delta v)/u} \tag{VI.1}$$

where m_0 is the initial mass, m is the mass remaining after a velocity increment Δv (in a practical case, gravity losses may occur so that the actual velocity increment is less than Δv), and $u = Ig$, where

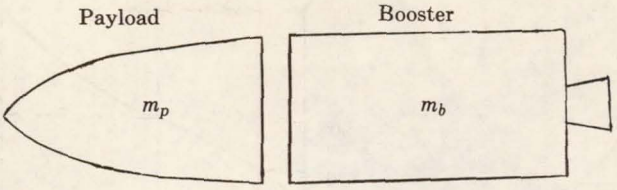


FIGURE 16. Booster with Payload

I is the specific impulse of the rocket (u is approximately the exhaust velocity of the gases in the rocket exhaust).

Consider the vehicle configuration of Figure 16. By (VI.1)

$$\frac{m_p + m_b}{m_p + \epsilon m_b} = K, \tag{VI.2}$$

where ϵm_b = burn out weight of booster.

Solve for m_b to get

$$m_b = \frac{K - 1}{1 - \epsilon K} m_p = r m_p. \tag{VI.3}$$

The required total weight is then

$$m_T = m_p + m_b = \frac{(1 - \epsilon) K}{1 - \epsilon K} m_p = R m_p. \tag{VI.4}$$

If it is required that a given velocity increment Δv be attained, one determines K from (VI.1), and then (VI.4) will show the initial weight of booster plus payload to give this increment of velocity to the payload.

b. *Direct Ascent to Moon and Return.* The path to be followed is shown in Figure 17. The configuration of the vehicle is shown

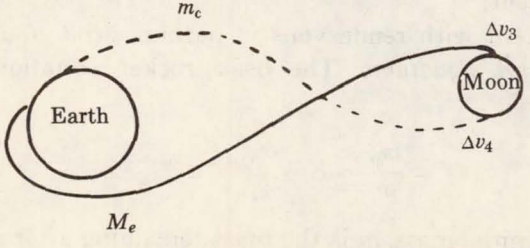


FIGURE 17. Direct Ascent Orbits

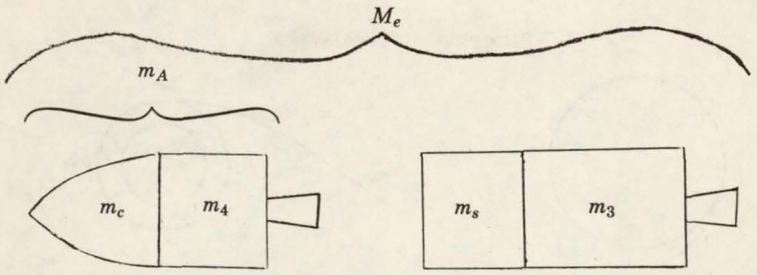


FIGURE 18. Direct Ascent Configuration

in Figure 18, where m_c is the mass of the capsule intended for Earth reentry, and m_s is the mass of supplies to be taken to the moon.

After the entire vehicle, with mass M_e, has been placed in an escape orbit directed toward the moon:

(i) Use m_3 to land on the moon; Δv_3 is the required velocity increment including gravity losses.

(ii) From the moon's surface, launch m_c to Earth reentry by means of m_4; Δv_4 is the velocity increment, including gravity losses and reserves.

We write:

$$m_A = R_4 m_c, \tag{VI.5}$$

$$M_e = R_3(m_A + m_s), \tag{VI.6}$$

where R_4 and R_3 would be determined by (VI.4). To determine the appropriate K's, values of Δv must be given. Values used in a typical mission study were

$$\Delta v_3 = 10{,}640 \text{ ft/sec},$$

$$\Delta v_4 = 10{,}330 \text{ ft/sec}.$$

c. *Lunar Rendezvous.* The path to be followed is shown in Figure 19. The configuration of the vehicle is shown in Figure 20, where m_c and m_s are as before.

After the entire vehicle, with mass m_e, has been placed in an escape orbit directed toward the moon:

(i) Use part of m_{12} to put the vehicle in a circular orbit about the moon; Δv_1 is the required velocity increment.

(ii) Detach m_L, leaving m_c and the rest of m_{12} in orbit.

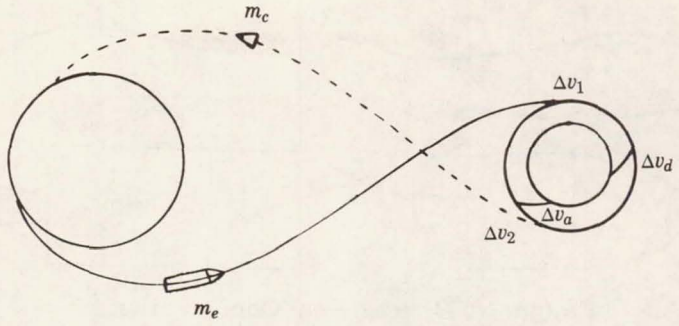


FIGURE 19. Lunar Rendezvous Orbits

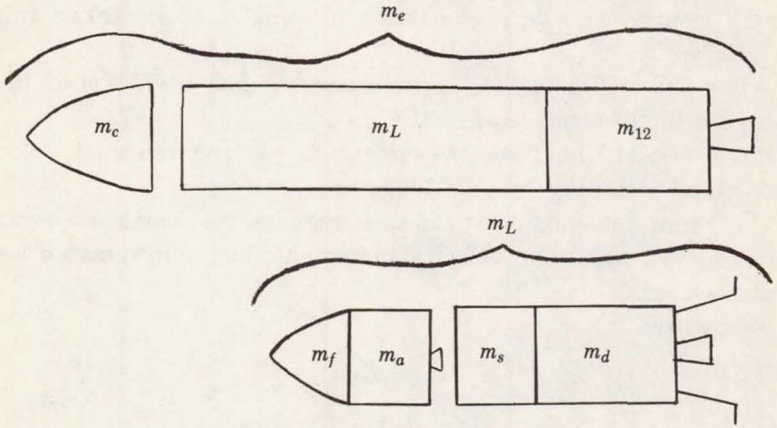


FIGURE 20. Lunar Rendezvous Configuration

(iii) Use m_d to descend to the moon; Δv_d is the required velocity increment, including gravity losses.

(iv) Use m_a to rendezvous with m_c and m_{12}; Δv_a is the required velocity increment, including gravity losses, rendezvous maneuvers, and reserves.

(v) Use the rest of m_{12} to return m_c to Earth reentry; Δv_2 is the velocity increment required.

We write:

$$m_a + m_f = R_a m_f, \tag{VI.7}$$

$$m_L = R_d (m_s + m_a + m_f). \tag{VI.8}$$

To get m_{12}, we denote by m_g the fuel required to produce Δv_1.

Then

$$\frac{m_{12} - m_g + m_c}{\epsilon m_{12} + m_c} = e^{\Delta v_2/u} = K_2,$$

$$\frac{m_{12} + m_c + m_L}{m_{12} - m_g + m_c + m_L} = e^{\Delta v_1/u} = K_1.$$

Solving these for m_{12} gives

$$m_{12} = \frac{K_1K_2 - 1}{1 - \epsilon K_1K_2} m_c + \frac{K_1 - 1}{1 - \epsilon K_1K_2} m_L. \tag{VI.9}$$

Finally

$$m_e = m_c + m_L + m_{12}. \tag{VI.10}$$

Values used in a typical mission study were:

$$\Delta v_a = 6800 \text{ ft/sec},$$

$$\Delta v_d = 6800 \text{ ft/sec},$$

$$\Delta v_2 = 3530 \text{ ft/sec},$$

$$\Delta v_1 = 3840 \text{ ft/sec}.$$

d. *Comparison of Missions.* For a direct ascent, we take $\epsilon_3 = 0.15$, and $\epsilon_4 = 0.12$. With the listed velocity increments, this gives

$$M_e = 10m_c + 2.7m_s. \tag{VI.11}$$

For lunar rendezvous, we take $\epsilon_{12} = 0.12$, and $\epsilon_a = \epsilon_d = 0.15$. With the listed velocity increments, this gives

$$m_e = 2.4m_c + 9.0m_f + 3.8m_s \tag{VI.12}$$

$$m_L = 5.5m_f + 2.4m_s. \tag{VI.13}$$

If we take $m_c = 13{,}000$ lb, $m_f = 3500$ lb, and $m_s = 0$ (for the early lunar missions) we find

$$M_e = 130{,}000 \text{ lbs},$$

$$m_e = 64{,}000,$$

$$m_L = 19{,}000.$$

These figures indicate the economy of a lunar rendezvous mission as compared to a direct ascent.

VII. **Rendezvous in interplanetary transfer.** Rendezvous concepts and paths for interplanetary flights are similar to those already discussed. However, special escape and capture maneuvers are needed in the near vicinity of the departure and destination planets.

Let V_p be the velocity around the sun needed by a space ship at earth distance from the sun if it is to reach the destination planet. If V_e is the earth's heliocentric velocity then the space ship must depart from earth with a relative velocity of V_∞ given by

$$V_\infty = V_p - V_e. \tag{VII.1}$$

We note that $V_e = 97{,}700$ fps.

If the ship starts from a circular "parking" orbit around the earth with velocity V_0, it must have this increased to a velocity V_E given by

$$V_E^2 = 2V_0^2 + V_\infty^2 \tag{VII.2}$$

if it is to proceed to a large distance from the earth with a residual relative velocity of V_∞. It will proceed along a hyperbolic path, as shown in Figure 21.

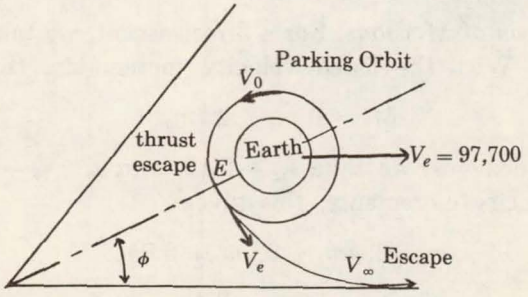


FIGURE 21. Hyperbolic Escape

The proper place at which to increase velocity from V_0 to V_E can be determined by noting that

$$\tan\phi = \frac{V_E V_\infty}{V_0^2}. \tag{VII.3}$$

At the destination planet a reverse procedure can be used.

Another type of escape maneuver uses a low thrust engine. By using continuous thrust for a long time, the space ship will traverse an ever ascending spiral until it escapes earth and takes up an elliptic orbit around the sun.

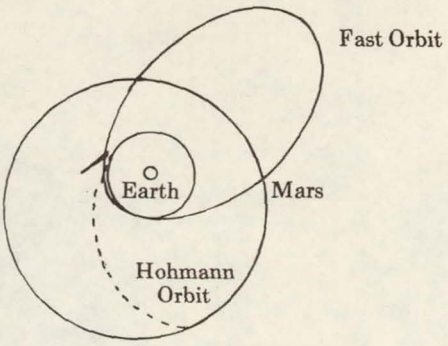


FIGURE 22. Orbits to Mass

The equations governing the motion during escape (by either method) are defined by the Lagrangian

$$L = T - V = \tfrac{1}{2}\, m(\dot{r}^2 + r^2\dot{\theta}^2) - \mu/r, \tag{VII.4}$$

and

$$\delta W = F_r \delta_r + F_\theta r \delta\theta. \tag{VII.5}$$

These yield the equations of motion for a thrusting escape:

$$\ddot{r} - r\dot{\theta}^2 + \frac{\mu}{r^2} = \frac{T}{m}\cos\beta, \quad r\ddot{\theta} + 2\dot{r}\dot{\theta} = \frac{T}{m}\sin\beta,$$

where T is thrust, m is vehicle mass, and β is the angle between the thrust vector and the local vertical.

The transfer path of minimum fuel expenditure is the so-called Hohmann transfer orbit; this is the dashed half ellipse shown in Figure 22. With a Hohmann orbit, the elapsed time can be quite large. There are faster orbits, of which one is shown in Figure 22. However, for them fuel expenditure can be excessive.

The characteristics for a Hohmann orbit and a selected fast orbit are given in Table 3.

TABLE 3. Orbit Characteristics

	Mars Stay Time	Travel Time	Total Time	$\sum \Delta V$
Hohmann	460 days	520 days	980 days	36,600 ft/sec
Fast Orbit	30	290	320	76,000

AERONAUTICAL RESEARCH ASSOCIATES OF PRINCETON, INC.
PRINCETON, NEW JERSEY

Index

Author Index for the Three Volumes

Roman numbers refer to pages on which a reference is made to a work of the author.

Italic numbers refer to pages on which a complete reference to a work by the author is given.

Boldface numbers indicate the first page of the articles in the three volumes. Roman numerals refer to Parts 1, 2, or 3, respectively.